A Datalog+/-Domain-Specific Durum Wheat Knowledge Base

Abdallah Arioua, Patrice Buche, Madalina Croitoru

A Datalog+/- Domain-Specific Durum Wheat Knowledge Base

Abdallah Arioua[1,2], Patrice Buche[2], Madalina Croitoru[1]

[1] LIRMM, University of Montpellier, France,
[2]IATE, INRA, France,

Abstract. We consider the application setting where a domain-specific knowledge base about Durum Wheat has been constructed by knowledge engineers who are not experts in the domain. This knowledge base is prone to inconsistencies and incompleteness. The goal of this work is to show how the state of the art knowledge representation formalism called *Datalog*± can be used to cope with such problems by (1) providing inconsistency-tolerant techniques to cope with inconsistency, and (2) providing an expressive logical language that allows representing incomplete knowledge.

1 Introduction

The Dur-Dur research project[1] aims at restructuring the Durum Wheat agrifood chain in France by reducing pesticide and fertilizer usage while providing a protein-rich Durum Wheat. The project relies on constructing a multidisciplinary knowledge base (involving all actors in the agrofood chain) which will be used as a reference for decision making. This knowledge base is collectively built by several knowledge engineers from different sites of the project. Due to various causes (errors in the factual information due to typos, erroneous databases / Excel files, incomplete facts, unspoken obvious information "everybody knows" etc.) the collectively built knowledge base (KB) is prone to *incompleteness and inconsistencies*. Incompleteness has many forms, in our case it reflects itself as a lack of precision and explicitness. For instance, an expert may say that the Durum Wheat is contaminated by a mycotoxin but he/she may, for some reasons, do not specify which mycotoxin. Inconsistency appears as logical contradictions due to the causes stated above. The problem is that in presence of inconsistencies the knowledge base becomes unreliable and not trustworthy, let alone the fact that reasoning under inconsistency is challenging for logical formalisms.

To solve the above mentioned problems, we propose in this paper a methodology of representing Durum Wheat knowledge in the logical framework of *Datalog*± [5, 10]. *Datalog*± is expressive enough to allow the representation of unknown individual in the knowledge base and cope with heterogeneous data as it allows for n-ary predicates. Moreover, *Datalog*± has an interesting equivalent relation with conceptual graphs [13], in fact any logical formula in *Datalog*± can be translated to a graphical representation,

[1] `http://www.agence-nationale-recherche.fr/?Projet=ANR-13-ALID-0002.`

which significantly helps experts in other domains in the process of knowledge acquisition. We present with detailed examples how this methodology is used to construct the Durum Wheat knowledge base in the French project Dur-Dur. The knowledge base is available online at `http://www.lirmm.fr/~arioua/dkb/` where the reader can find downloadable materials.

2 The Logical Language *Datalog*$\pm$

There are two major approaches in the knowledge representation community: Description Logics (DL) (such as $\mathcal{EL}$ [2] and DL-Lite [6] families) and rule-based languages (such as *Datalog*$\pm$ language[10, 5], a generalization of Datalog that allows for existentially quantified variables in rule's head). Despite its undecidability when answering conjunctive queries, different decidable fragments of *Datalog*$\pm$ are studied in the literature [4]. These fragments generalize the above mentioned DL families and overcome their limitations by allowing any predicate arity as well as cyclic structures.

The *Datalog*$\pm$ corresponds to *the positive existential* conjunctive fragment of first-order logic, which is composed of formulas built with the connectors $(\wedge, \rightarrow)$ and the quantifiers $(\exists, \forall)$, with constants but no function symbol.

An *atom* is of the form $p(t_1, \ldots, t_k)$ where p is a predicate of arity k and the t_i are terms, i.e., variables or constants (we use vectors , e.g. $\overrightarrow{x}$, to denote a sequence of variables). A finite set of atoms F is called an *atomset* (a *fact*), we denote by $terms(F)$ (resp. $vars(F)$) the set of terms (resp. variables) that occur in F. A *homomorphism* π from two atomsets A_1 to A_2 is a substitution of $vars(A_1)$ by $terms(A_2)$ such that $\pi(A_1) \subseteq A_2$. An *existential rule* (or a rule) is of the form $R = \forall \overrightarrow{x} \forall \overrightarrow{y} (B \rightarrow \exists \overrightarrow{z} H)$, where B and H are conjunctions of atoms, with $vars(B) = \overrightarrow{x} \cup \overrightarrow{y}$, and $vars(H) = \overrightarrow{x} \cup \overrightarrow{z}$. B and H are respectively called the *body* and the *head* of R. Chase is the mechanism by which one deduce new facts by rule application on the initial set of facts $\mathcal{F}$. We denote by $\texttt{Cl}_{\mathcal{R}}(\mathcal{F})$ the set of all facts that can be deduced from $\mathcal{F}$ by a set of rules $\mathcal{R}$. A *knowledge base* $\mathcal{K} = (\mathcal{F}, \mathcal{R}, \mathcal{N})$ is composed of a finite set of facts $\mathcal{F}$, rules $\mathcal{R}$ and negative constraints $\mathcal{N}$ (i.e. a rule whose head is set to $\bot$). A *Boolean conjunctive query* (BCQ or query in the following) has the form of a fact. We say a query Q is entailed from $\mathcal{K}$ iff $\texttt{Cl}_{\mathcal{R}}(\mathcal{F}) \models Q$. We say $\mathcal{K}$ is inconsistent iff $\texttt{Cl}_{\mathcal{R}}(\mathcal{F}) \models \bot$.

3 The Durum Wheat Knowledge Base

The Durum Wheat knowledge base has been constructed within the French National Project DUR-DUR. The goal of this knowledge base is to integrate scientific knowledge acquired from different tasks during the project to redesign the durum wheat chain. The Dur-Dur project suggests developing a systematic approach to investigate issues related to the management of the nitrogen, energy and contaminants, to guarantee a global quality of products throughout the production and the processing chain. Started in 2014 and planned over 4 years, the project aims at integrating the 3 dimensions of the sustainability (environmental, economic, and social), at 4 levels of investigation (4 tasks) with a complementary task (task 5). Figure 1 depicts the different tasks of the project where the fifth task's central role is to integrate knowledge from different tasks. The

Durum Wheat knowledge base is the product of the fifth Task. It will be used in many computational tasks, notably analyzing and comparing the alternative innovative technical itineraries proposed in the project to reduce the use of chemical inputs (nitrogen fertilizers and pesticides). The knowledge base represents domain-specific knowledge about Agronomy. It is composed of four main parts:

- **Vocabulary:** it contains knowledge about concepts and relations.
- **Rules:** they represent rules that encode generic knowledge.
- **Negative constraints:** this part contains constraints about crops and Agronomy-related constraints.
- **Facts:** this part contains factual knowledge about Agronomy-related subjects (fertilizers, pesticides, diseases,etc.).

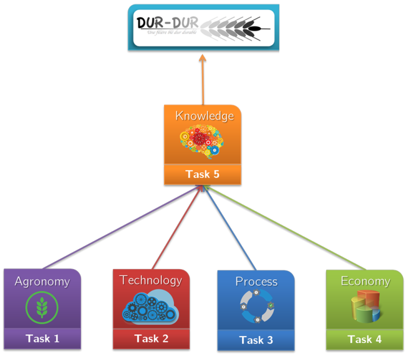


Fig. 1: The different tasks of the Dur-Dur project. The knowledge tasks aims at integrating multidisciplinary knowledge from other tasks.

In the next section we start by highlighting the guidelines which were followed to author the knowledge base (Subsection 3.1) then we turn to the the internal structure or the architecture of the knowledge base including the *vocabulary*, the *rule-base* alongside with the *constraints* and the *factual knowledge* (Subsection 3.2).

3.1 The Authoring

A multidisciplinary process of knowledge acquisition and representation was deployed to author the knowledge base. We used technical reports to define the scope of the knowledge base and the relevant concepts of our vocabulary. Taking into account the recommendation of [15], we followed three steps *specification, conceptualization and formalization* to build the knowledge base.

Specification. The scope of the Durum Wheat knowledge base has been defined by exclusively focusing on *Durum Wheat Sustainability* management. The goal is improving Durum Wheat sustainability in France and reduce the use of nitrogen fertilizers and

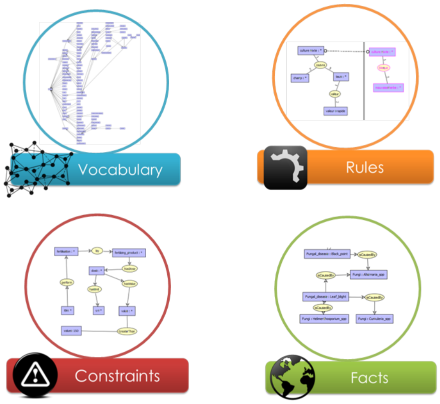


Fig. 2: An overview of the Durum Wheat knowledge base. The circles contains knowledge examples represented in the conceptual graph framework.

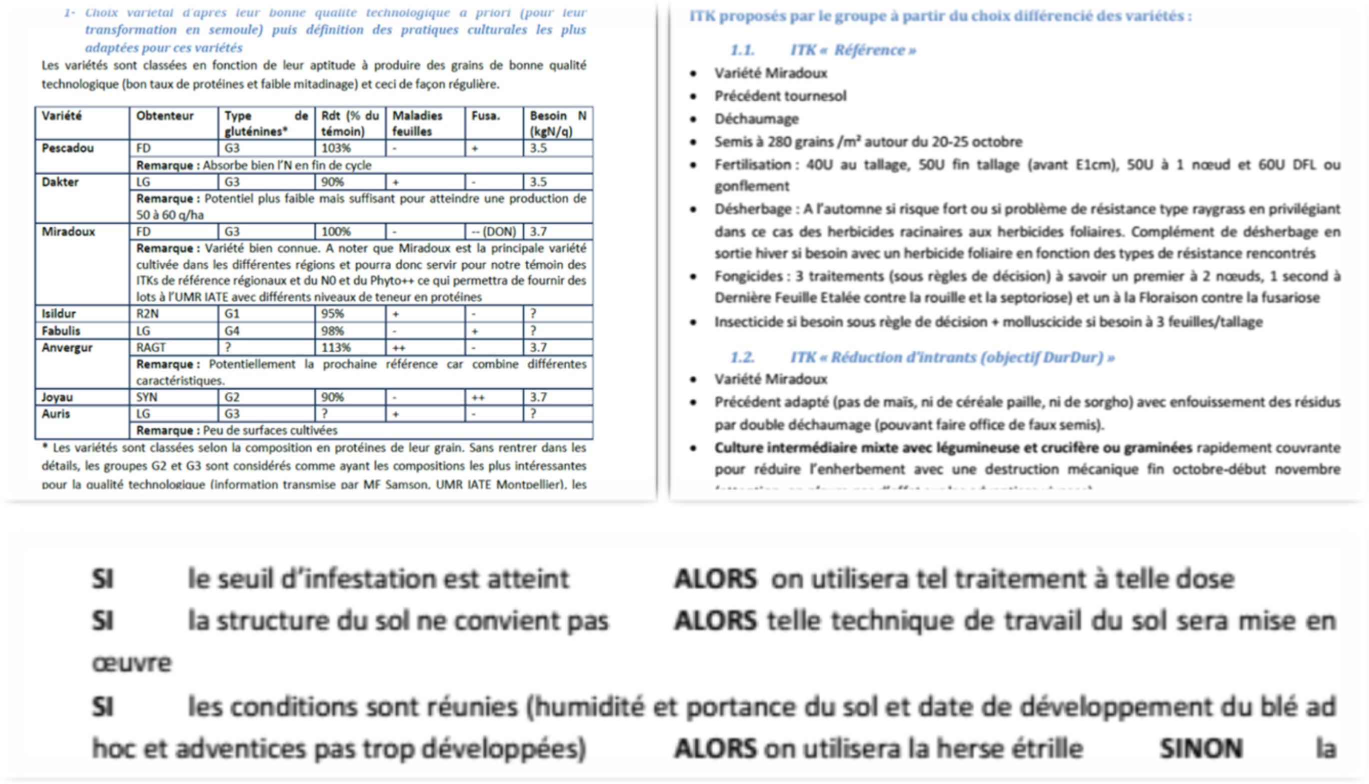
pesticides and optimize energy consumption using a systematic approach that makes use of innovative technical itineraries. The contribution of the knowledge base lays in offering an expressive way of representing domain-knowledge.

Conceptualization. The concepts and the relations among them alongside to rules, facts and constraints have been defined and collected from technical reports (see Figure 3) and online materials (see [1]). It is worth mentioning that in the vocabulary part we have built on the vocabulary of Agropedia indica ([12]) with an increase (and modification) in content that approximates 60%.[2]

Formalization. Since understanding logical formulas is quite difficult for experts who are not familiar with KRR formalism we have chosen a graphical framework (conceptual graphs; [13, 14]) to author the knowledge base. Moreover, the conceptual graphs (CGs) made it easy for the Agronomy experts to understand the content of the knowledge base. Furthermore, CGs enjoy the same expressive power as *Datalog*±. In fact, it is an equivalent formalism of *Datalog*± as shown in [7]. Therefore, our choice was to choose CGs for knowledge acquisition and *Datalog*± as a framework for reasoning. For CGs, we used *CoGui 1.6b* which is an IDE for representing and reasoning with CGs.[3] We shall explain in-depth in Section 3.2 the graphical and logical representation for each part of the knowledge base. The facts within the knowledge base are exported to an RDF/XML format whereas the vocabulary, rules and constraints are exported as DLGP format (**D**ata**L**o**G** **P**lus; [8]). The vocabulary of the knowledge base contains 279 concepts and 116 relations, the rule-base contains 23 rules and the constraints part contains 25 constraints. The factual part has around 900 atoms. The knowledge base is available online at `http://www.lirmm.fr/~arioua/dkb/`.

[2] `http://www.agropedia.net`.

[3] `http://www.lirmm.fr/cogui/`.



1- Choix variétal d'après leur bonne qualité technologique a priori (pour leur transformation en semoule) puis définition des pratiques culturales les plus adaptées pour ces variétés

Les variétés sont classées en fonction de leur aptitude à produire des grains de bonne qualité technologique (bon taux de protéines et faible mitadinage) et ceci de façon régulière.

Variété	Obtenteur	Type de gluténines*	Rdt (% du témoin)	Maladies feuilles	Fusa.	Besoin N (kgN/q)
Pescadou	FD	G3	103%	-	+	3.5
	Remarque : Absorbe bien l'N en fin de cycle					
Dakter	LG	G3	90%	+	-	3.5
	Remarque : Potentiel plus faible mais suffisant pour atteindre une production de 50 à 60 q/ha					
Miradoux	FD	G3	100%	-	-- (DON)	3.7
	Remarque : Variété bien connue. A noter que Miradoux est la principale variété cultivée dans les différentes régions et pourra donc servir pour notre témoin des ITKs de référence régionaux et du N0 et du Phyto++ ce qui permettra de fournir des lots à l'UMR IATE avec différents niveaux de teneur en protéines					
Isildur	R2N	G1	95%	+	-	?
Fabulis	LG	G4	98%	-	+	?
Anvergur	RAGT	?	113%	++	-	3.7
	Remarque : Potentiellement la prochaine référence car combine différentes caractéristiques.					
Joyau	SYN	G2	90%	-	++	3.7
Auris	LG	G3	?	+	-	?
	Remarque : Peu de surfaces cultivées					

* Les variétés sont classées selon la composition en protéines de leur grain. Sans rentrer dans les détails, les groupes G2 et G3 sont considérés comme ayant les compositions les plus intéressantes pour la qualité technologique (information transmise par MF Samson, UMR IATE Montpellier), les

ITK proposés par le groupe à partir du choix différencié des variétés :

1.1. ITK « Référence »

- Variété Miradoux
- Précédent tournesol
- Déchaumage
- Semis à 280 grains /m² autour du 20-25 octobre
- Fertilisation : 40U au tallage, 50U fin tallage (avant E1cm), 50U à 1 nœud et 60U DFL ou gonflement
- Désherbage : A l'automne si risque fort ou si problème de résistance type raygrass en privilégiant dans ce cas des herbicides racinaires aux herbicides foliaires. Complément de désherbage en sortie hiver si besoin avec un herbicide foliaire en fonction des types de résistance rencontrés
- Fongicides : 3 traitements (sous règles de décision) à savoir un premier à 2 nœuds, 1 second à Dernière Feuille Etalée contre la rouille et la septoriose) et un à la Floraison contre la fusariose
- Insecticide si besoin sous règle de décision + molluscicide si besoin à 3 feuilles/tallage

1.2. ITK « Réduction d'intrants (objectif DurDur) »

- Variété Miradoux
- Précédent adapté (pas de maïs, ni de céréale paille, ni de sorgho) avec enfouissement des résidus par double déchaumage (pouvant faire office de faux semis).
- **Culture intermédiaire mixte avec légumineuse et crucifère ou graminées** rapidement couvrante pour réduire l'enherbement avec une destruction mécanique fin octobre-début novembre

SI le seuil d'infestation est atteint **ALORS** on utilisera tel traitement à telle dose

SI la structure du sol ne convient pas **ALORS** telle technique de travail du sol sera mise en œuvre

SI les conditions sont réunies (humidité et portance du sol et date de développement du blé ad hoc et adventices pas trop développées) **ALORS** on utilisera la herse étrille **SINON** la

Fig. 3: Some snapshots of the technical reports.

3.2 The Structure

As depicted in Figure 2 the knowledge base is composed of four parts. It is worth mentioning that on the logical level the vocabulary and the rule-base are the same. However, we adapt here the the Semantic web notation and we differentiate between them. Therefore, we distinguish between those rules that express logical consequences (in the rule-base) and those that encode generalizations and classes inclusions (in the vocabulary).

The vocabulary The vocabulary represents an explicit specification of the terms and concepts used in Agronomy. The vocabulary is composed of two parts: (1) concept types hierarchy and (2) relation types hierarchy.

1. **Concept types hierarchy:** concepts are organized within a hierarchy as super-concepts and sub-concepts. For instance, the concept disease and its sub-concepts (e.g. viral disease, fungal disease, etc.), types of pesticides (e.g. herbicide, insecticide, fungicide) are all of organized in a hierarchy.
2. **Relation types hierarchy:** in CGs the concepts are related by relationships. Since concepts are divided into super-concepts and sub-concepts, relationships are divided in the same way. In the relation types hierarchy we find super-relations and sub-relations. For instance, the relation "useSowingProcess" which relates the *seeding and sowing* production step with the process of *sowing* (which is a super-concept of *broadcasting*, *behind plough* and a sub-concept of *process*). This relation is a sub-relation of the super-relation "useProcess" that relates any production step with any process.

In CGs the hierarchy of concept types is represented as in the upper graph of Figure 4. Rectangles represent concepts and the arrow represents the generalization between them where the source of the arrow is the sub-concept and the target of the arrow is super-concept. In the relation types hierarchy (the lower graph), the circles are the relations and the arrows are generalizations.

To better illustrate the relation between existential rules and CGs let us take an example that shows the transformation of some part of the graphs of Figure 4 to their logical form.

Example 1. The left-most part of the concept types hierarchy that indicates that "Viral disease is a disease" is represented logically by a rule as follows:

– $\forall x(Viral\_disease(x) \rightarrow Disease(x))$.

The part of the relation types hierarchy that indicates that "Using Herbicide is using Pesticide" is represented logically by a rule as follows:

– $\forall x, y(useHerbicide(x, y) \rightarrow usePesticide(x, y))$.

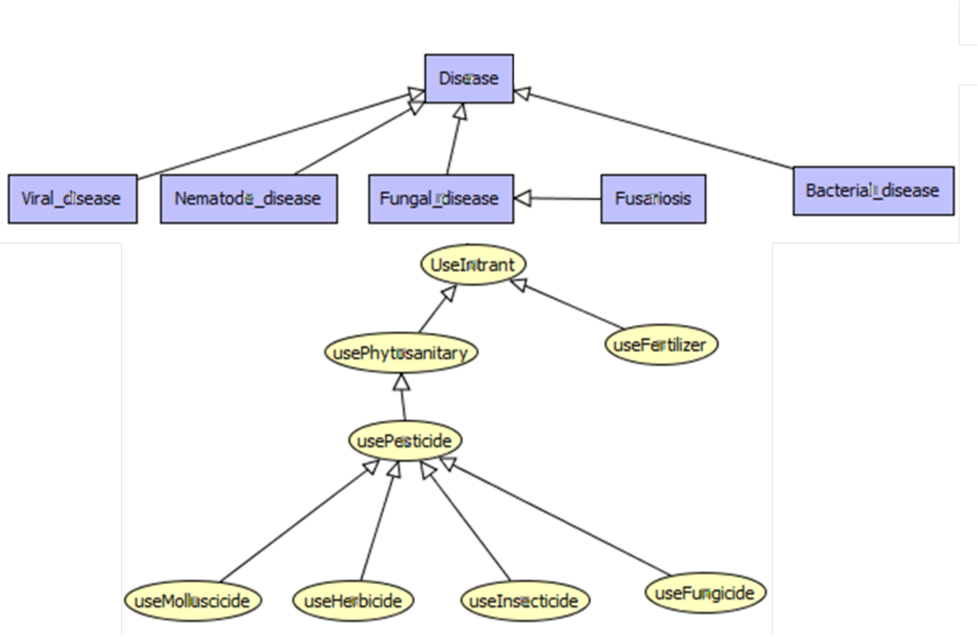


Fig. 4: Concept and relation types hierarchy.

The rule-base Rules in the rule-base encode general-purpose domain-specific knowledge. For instance, consider the following rules:

(a) If a Durum Wheat x has a fusariosis disease y then there exists a mycotoxin z that has contaminated the Durum Wheat x.
$\forall x, y \exists z (Durum\_wheat(x) \wedge hasDisease(x, y) \wedge Fusariosis(y) \rightarrow isContaminatedBy(x, z) \wedge Mycotoxin(z))$

(b) If the soil is rich of organic matters and it contains seeds of weed then these seeds will develop in this soil.
$\forall x, y, z, w (Soil(x) \wedge Organic\_matter(y) \wedge richOf(x, y) \wedge contains(x, z) \wedge Seed(z) \wedge seedOf(z, w) \wedge Weed(w) \rightarrow developIn(w, x))$

The mycotoxin z is unknown (it could be Aflatoxins, Deoxynivalenol, etc.) but still the information that “there is necessarily a mycotoxin” is present, which is an important information when it comes to risk management where a possible contamination by any mycotoxin is taken to be critical. Moreover, the importance of such representation manifests also in helping knowledge elicitation where the knowledge base can make use of incomplete information and then be updated incrementally by identifying the existential variables.

In conceptual graphs the rule (b) is depicted in Figure 5. In a rule, the rectangles are called *concept nodes* and the circles are called *relation nodes*. A concept node has a *concept type* and a *marker* which can be either an *individual marker* (constant) or a *generic marker* (a variable denoted as *). For instance, the concept $richOf$ has a generic marker (*) which represents a variable. If the marker were an individual marker we should have found a constant name like *Nitrogen*. The relation nodes are predicates that relate different concepts. A rule in conceptual graphs is composed of two parts, a hypothesis (left) and a conclusion (right). The dashed lines link those concepts that share the same variables (called frontier variables). That means, variables that appear in the hypothesis and in the conclusion. In the rule (b), the concept $weed$ in the hypothesis part shares the same variable with the concept $weed$ in the conclusion.

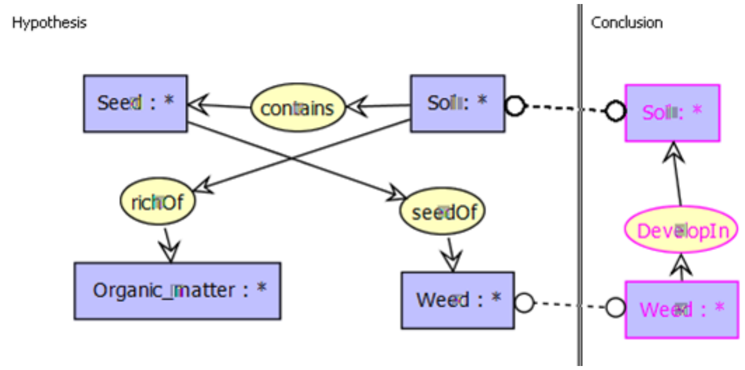


Fig. 5: The rule (b) in the CGs framework.

As we said earlier, certain classes of *Datalog*± render the inference undecidable. However, there are some classes that ensure decidability. Most notably, FUS (Finite Unification Set) and FES (Finite Expansion Set) classes. The online tool *Kiabora* deploys syntactical and semantic analysis on any set of rules written in the DLGP format.[4] The tool, for a given rule-base, classifies all the rules with respect to the known classes. From the analysis we found that our rule-base lays within the decidable classes. Specifically, FUS and FES.

The negative constraints Representing what cannot be allowed within certain domain of interest is called *negative constraint* (or constraint). Consider the following negative constraint:

(c) $\forall x, y, z(Soil(x) \wedge Maize(y) \wedge Durum\_wheat(z) \wedge hasPrecedent(x, y)$
$\wedge\, isCultivatedOn(z, x) \rightarrow \bot).$

This negative constraint forbids using Maize as a precedent on a soil if we want to cultivate Durum Wheat on this soil. Figure 6 represents the CGs representation of this negative constraint. Besides this type of constraints we have the banned types constraints. These are particular forms of constraints that express concept disjointness. For

[4] Kiabora 0.1 website: `http://www.lirmm.fr/~mugnier/graphik/kiabora/`, see [8] for a detailed explanation.

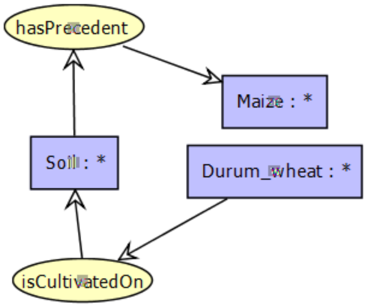


Fig. 6: The negative constraint (c) in the CGs framework.

instance, a soil x cannot be a disease, $\forall x(Soil(x) \wedge disease(x) \rightarrow \bot)$. In the Durum Wheat knowledge base all concepts are disjoint except those concepts which have a generalization/specialization relations among them.

The factual knowledge In the Durum Wheat knowledge base the factual part represents domain-specific knowledge. This knowledge is divided into two parts: (1) general factual knowledge and (2) knowledge about different technical itineraries. According to [11] a *technical itinerary* is a "*logical organized course of technical actions* applied to a cropped species".

General factual knowledge is the part of the knowledge base that represents general facts about the domain, for instance, *Miradoux* is a variety of Durum Wheat or the fungal disease *Fusarium Flag smut* is cause by, among other causes, the fungi *Urocyctis agropyri* of the family Fusarium. The following is an example of a set of facts. Recall that commas are interpreted as conjunctions.

(d) $\{Fungal\_disease(Flag\_smut), isCausedBy(Fusarium\_ear\_blight,$ $Urocyctis\_agropyri), fungi(Urocyctis\_agropyri)\}$.

Here we have the relation $isCausedBy$ instantiated on the individuals $Flag\_smut$ and $Urocyctis\_agropyri$. The former is a fungal disease as stated by the concept $Fungal\_disease$ and the latter is a fungi. Figure7 depicts the set of facts in the CGs framework.

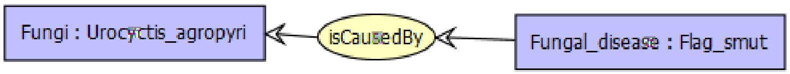


Fig. 7: The set of facts (d) about fungal diseases and fungus.

The second part of the factual knowledge part are those facts about the technical itineraries. In what follows we give a real-world example of a well-known technical itinerary in France.

Example 2. This example represents the reference technical itinerary in France which is followed by farmers to cultivate their fields.

"The variety to be seeded in the soil is Miradoux, the culture precedent is sunflower. The soil is prepared by means of harrowing. The seeding is done with density of 280 grains/m^2. Fertilization is to be performed at the growing stage when the tiller begins with dose 40u and 50u at the end of the tiller."

This technical itinerary is a set of facts, e.g. "variety is Miradoux", "Fertilization is to be performed at the growing stage", etc. However, not any any set of facts. Particularly, it is a precise set of describing facts. Actually, any ITK (according to the studied reports) should precisely account for the following steps:

1. Variety to bee seeded.
2. Date of seeding alongside the density.
3. Cultural precedent.
4. Inter-cropping techniques.
5. Soil preparation method.
6. Disease management method.
7. Weed management method.
8. Insect control method.

Thus a technical itinerary should be mainly composed of these describing facts. The fowling is a snippet of the technical itinerary described in Example 2.

$$\mathcal{F}_{ITK} = \left\{ \begin{array}{ll} Soil(Soil_1) & Durum\_wheat(D1) \\ isOfVariety(D1, Miradoux) & Variety(Miradoux) \\ isCultivatedOn(D1, Soil_1) & Seeding\_and\_sowing(Seeding_1) \\ Seed(Seed_1) & useSeed(Seeding1, Seed_1) \\ seedOf(Seed_1, Durum_w 1) & isAppliedOn(Seeding_1, Soil_1) \\ withDensity(Seeding_1, Density_1) & Density(Density_1) \\ Unit(grain\_mm) & hasValue(280) \\ Value(280) & \end{array} \right.$$

What has been presented so far is the Durum Wheat knowledge base we have constructed manually within the project, which can be seen as a contribution on itself. since what is mainly proposed by researchers in this field are ontologies. Besides that, our knowledge base provides querying facilities not only on the ontological layer but also on the factual layer where real knowledge about the domain is represented in form of facts.

In the Durum Wheat knowledge base each technical itinerary is stored separately from the other technical itineraries (we have three in total). On can query them all together or separately.

3.3 Reasoning

The first and foremost reason to acquire knowledge and store it in knowledge bases is to provide querying facilities for the end-user. Like in classical database systems, in *Datalog*$\pm$ the main reasoning task is *query answering*. The main and important difference is that in our case the querying is enriched by a rule-base layer. Thus the reasoner takes into account the domain-knowledge represented within the rules while querying.

Formally, a conjunctive query has the form of a fact but with possibly free variables. For instance $Q(x) = Fungal\_disease(x) \wedge isCausedBy(x, culmorum)$ is a conjunctive query that looks for "the fungal disease that is caused by $culmorum$".

In order to perform reasoning in forward chaining in presence of rules, the reasoner applies all the rules in the rule-base on the set of facts in the factual part then query the knowledge in a classical manner.Given a set of facts $\mathcal{F}$ and a set of rules $\mathcal{R}$ This means that the chase computes all deducible knowledge of $\mathcal{F}$ by the application of all the rules of $\mathcal{R}$ on all the facts on $\mathcal{F}$ until no rule will be applicable. This process is also called *saturation*. Note that if the closure of a set of facts $\mathcal{F}$ is the same as $\mathcal{F}$, i.e. $\mathtt{Cl}_{\mathcal{R}}(\mathcal{F}) = \mathcal{F}$, then we say that $\mathcal{F}$ is closed under the application of rules (or deductively closed). A query Q has an answer within a knowledge base $\mathcal{K}$ iff $\mathtt{Cl}_{\mathcal{R}}(\mathcal{F}) \models Q$ where $\models$ refers to the usual first-order entailment.

Example 3. Consider the following knowledge base $\mathcal{K}$:
$\mathcal{F} = \{D(a), S(b)\}$, $\mathcal{R} = \{\forall x(D(x) \rightarrow C(x)), \forall x, y(S(x) \wedge C(y) \rightarrow M(x, y))\}$, $\mathcal{N} = \emptyset$. The closure is $\mathtt{Cl}_{\mathcal{R}}(\mathcal{F}) = \{D(a), S(b), C(a), M(a, b)\}$.

It may happens that the set facts $\mathcal{F}$ contains contradictory knowledge (i.e. inconsistencies). We say that a set of facts is inconsistent iff $\mathtt{Cl}_{\mathcal{R}}(\mathcal{F})$ triggers a negative constraint. The solution [9] is to construct maximal (with respect to set inclusion) consistent subsets of $\mathcal{F}$. Such subsets are called *repairs* and denoted by $\mathcal{R}epair(\mathcal{K})$. They actually represent possible distribution of facts to restore consistency. Once the repairs are computed, different semantics can be used for query answering over the knowledge base.

Example 4. Consider: $\mathcal{F} = \{D(a), S(b), P(c)\}$, $\mathcal{R} = \{\forall x(D(x) \rightarrow C(x))\}$, $\mathcal{N} = \{\forall x, y(S(x) \wedge C(x)) \rightarrow \perp\}$. Then the negative constraint will be triggered after the application of the rule which infers $C(a)$. Therefore our repairs would be $\mathcal{A}_1 = \{D(a), P(c)\}$ and $\mathcal{A}_2 = \{S(b), P(c)\}$ and $\mathcal{R}epair(\mathcal{K}) = \{A_1, A_2\}$. While $\mathtt{Cl}_{\mathcal{R}}(\mathcal{A}_1) = \{D(a), C(a), P(c)\}$ and $\mathtt{Cl}_{\mathcal{R}}(\mathcal{A}_2) = \mathcal{A}_2$.

After repairing the knowledge base we can query it using different semantics. The most common semantics is to query the intersection of all repairs. This is a cautious strategy because the intersection is practically those facts which are not involved in any inconsistency.

4 Conclusion

In this paper we have presented a general methodology to build Durum Wheat knowledge bases within the logical language *Datalog*$\pm$. We presented detailed examples and a real-world Durum Wheat knowledge base which has been built within the French national project Dur-Dur. The expressiveness of *Datalog*± lays in its ability to deal with incompleteness and inconsistency. Moreover, it has an interesting relation with Conceptual Graphs which makes it easy to non-experts to manipulate and understand logical formulae. In addition, DLGP format (**D**ata**L**o**G** **P**lus; [8]) can be translated to semantic web languages as OWL/RDFS using COGui or GRAAL framework [3].

References

1. Arvalis. Les fiches arvalis: Varits produits accidents en grandes cultures. Online materials published by ARVALIS Plant Institue., 2015. Available at http://www.fiches.arvalis-infos.fr/, 19-03-2015.
2. F. Baader, S. Brandt, and C. Lutz. Pushing the el envelope. In *Proc. of IJCAI 2005*, 2005.
3. J. Baget, M. Leclère, M. Mugnier, S. Rocher, and C. Sipieter. Graal: A toolkit for query answering with existential rules. In N. Bassiliades, G. Gottlob, F. Sadri, A. Paschke, and D. Roman, editors, *Rule Technologies: Foundations, Tools, and Applications - 9th International Symposium, RuleML 2015, Berlin, Germany, August 2-5, 2015, Proceedings*, volume 9202 of *Lecture Notes in Computer Science*, pages 328–344. Springer, 2015.
4. J.-F. Baget, M.-L. Mugnier, S. Rudolph, and M. Thomazo. Walking the complexity lines for generalized guarded existential rules. In *Proc. of IJCAI'11*, pages 712–717, 2011.
5. A. Calì, G. Gottlob, and T. Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. *J. Web Sem.*, 14:57–83, 2012.
6. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The dl-lite family. *J. Autom. Reasoning*, 39(3):385–429, 2007.
7. M. Chein and M. Mugnier. *Graph-based Knowledge Representation - Computational Foundations of Conceptual Graphs*. Advanced Information and Knowledge Processing. Springer, 2009.
8. M. Leclère, M.-L. Mugnier, and S. Rocher. Kiabora: an analyzer of existential rule bases. In *Web Reasoning and Rule Systems*, pages 241–246. Springer, 2013.
9. D. Lembo, M. Lenzerini, R. Rosati, M. Ruzzi, and D. F. Savo. Inconsistency-tolerant semantics for description logics. In *Proceedings of RR'10*, pages 103–117. Springer-Verlag, 2010.
10. M. Mugnier and M. Thomazo. An introduction to ontology-based query answering with existential rules. In M. Koubarakis, G. B. Stamou, G. Stoilos, I. Horrocks, P. G. Kolaitis, G. Lausen, and G. Weikum, editors, *Reasoning Web. Reasoning on the Web in the Big Data Era - 10th International Summer School 2014, Athens, Greece, September 8-13, 2014. Proceedings*, volume 8714 of *Lecture Notes in Computer Science*, pages 245–278. Springer, 2014.
11. M. Sebillotte. Itineraires techniques et evolution de la pensee agronomique. *Comptes Rendus des Séances de l'Académie d'Agriculture de France*, 1978.
12. M. Sini and V. Yadav. *Building Knowledge Models for Agropedia Indica v 1.0 Requirements, Guidelines, Suggestions*, 2009. Available at http://agropedia.iitk.ac.in/km_guidlines.pdf, visited 19-03-2015.
13. J. F. Sowa. Conceptual graphs for a data base interface. *IBM Journal of Research and Development*, 20(4):336–357, 1976.
14. J. F. Sowa. Conceptual structures: information processing in mind and machine. 1983.
15. A. Thunkijjanukij, A. Kawtrakul, S. Panichsakpatana, and U. Veesommai. Lesson learned for ontology construction with thai rice case study. In *World conference on agricultural information and IT, IAALD AFITA WCCA 2008*, pages 495–502, 2008.