Weak Subsumption in the $\mathcal{EL}$-Description Logic with Refreshing Variables (Extended Abstract)

Théo Ducros, Marinette Bouet, and Farouk Toumani

LIMOS, 1 rue de la Chébarde, Aubière 63178 France firstname.name@limos.fr

Description Logics with Refreshing Variables. Concepts with variables (also called patterns) have been introduced in description logics since the mid-nineties [1, 10, 12] and led to a highly interesting research stream on the so-called non-standard reasoning, specifically matching [6, 4] and unification [3]. As an example to briefly recall these notions, consider the following pattern: $P_1 \equiv Y \sqcap \exists x.University$. Here, the variable x, called a *role variable*, takes its values from a set of possible atomic role names while the variable Y, called *concept variable*, takes its values among all possible concept descriptions. The concept description $FemaleStudent \equiv Person \sqcap Female \sqcap \exists studyIn.University$ *matches* the pattern P_1. Indeed, if we consider a variable substitution σ that assigns the role *studyIn* to the variable x (i.e., $\sigma(x) = studyIn$) and assigns the concept description $Person \sqcap Female$ to the variable Y (i.e., $\sigma(Y) = Person \sqcap Female$), the concept $\sigma(P_1) \equiv \sigma(Y) \sqcap \exists \sigma(x).University$ obtained by replacing the variables x and Y by their respective values given by the substitution σ, is equivalent to the concept $FemaleStudent$. Hence, given a description C and a pattern P, a matching problem modulo equivalence (respectively, modulo subsumption) asks then whether there is a variable substitution σ such that $C \equiv \sigma(P)$ (respectively, $C \sqsubseteq \sigma(P)$). Unification extends matching to the case where C is itself a pattern.

Matching and unification have shown to be useful to filter unimportant aspects of large concepts [11], to detect redundancies in knowledge bases [7] and to support integration of knowledge bases [9]. Most of existing works regarding these two inference tasks deal with concept variables. Matching between $\mathcal{FL}_0$ terms with concept variables has been shown polynomial [1] while considering general TBox induces a blow up in complexity leading to EXPTIME [4]. On the other hand, matching in $\mathcal{EL}$ in presence or not of TBox is NP-complete. Unification in $\mathcal{FL}_0$ is EXPTIME-complete [7] while it is NP-complete in $\mathcal{EL}$ [5] and stays in NP for $\mathcal{EL}$ in presence of cycle-restricted TBoxes [3]. Role variables have been studied in [10, 11] for $\mathcal{FL}_0$. In [11], it has been shown that matching between $\mathcal{FL}_0$-terms with both concept and role variables is NP-Complete.

Our work explores a new class of variables, called *refreshing variables*. More precisely, we support cyclic pattern definitions and consider a new semantics for variables, called *refreshing semantics* in contrast to the classical (*non refreshing*) semantics used in the literature. The main difference between these two kinds of semantics lies in the valuation of variables that appear in the scope of a terminological cycle. A classical semantics requires to have a unique valuation of such variables while a refreshing semantics enables to assign different values for the same variable for each *unfolding* of a cycle. To illustrate the refreshing semantics, let us consider the following pattern $Academic \equiv Person \sqcap \exists x.University \sqcap \exists y.Academic$. According to this pattern, an $Academic$ is defined as a person having *certain relationships* with a $University$ and an $Academic$. Consider now the following partial unfolding of this pattern : $Academic \equiv Person \sqcap \exists x_1.University \sqcap \exists y_1.(\exists x_2.University \sqcap \exists y_2.Academic)$ where indices are used to distinguish different occurrences of the same variable during the unfolding process. In classical semantics, a variable substitution σ assigns a unique value to all the occurrences of a given variable x (i.e., $\sigma(x_i) = \sigma(x_j), \forall i, j$). In contrast, refreshing semantics permits substitutions that may assign different values to different occurrences of a given variable (i.e., we may have $\sigma(x_i) \neq \sigma(x_j)$ for $i \neq j$). Intuitively, refreshing semantics enables a variable to be refreshed at each cycle of the unfolding process.

Considering matching or unification in the context of a refreshing semantics leads to the following practical benefits: *(i)* some matching (or unification) problems that are unsolvable under the classical semantics have a solution under the refreshing semantics, and *(ii)* for some matching (or unification) problems, the refreshing semantics provides solutions that are more general w.r.t. the subsumption relation and hence are of practical interest [10]. We illustrate these issues by considering the pattern $Academic$ and the following $\mathcal{EL}$-Tbox $\mathcal{T}$:

$PhDStudent$	$\equiv$	$Person \sqcap \exists studyIn.University \sqcap \exists supervisedBy.Doctor$
$Doctor$	$\equiv$	$Person \sqcap \exists getPhDIn.University \sqcap \exists formerly.PhDStudent$
$FrenchUniv$	$\equiv$	$University \sqcap \exists located.France$
$FrenchPhDStudent$	$\equiv$	$Person \sqcap \exists studyIn.FrenchUniv \sqcap \exists supervisedBy.Doctor$
$FrenchDoctor$	$\equiv$	$Person \sqcap \exists getPhDIn.FrenchUniv \sqcap$ $\exists formerly.FrenchPhDStudent$

We consider the following classical matching problem w.r.t a Tbox in the setting of the $\mathcal{EL}$ description logic: $Doctor \equiv^?_{\mathcal{T}} Academic$. In this setting, x, y are non refreshing role variables and hence they have a finite number of possible values (i.e., the role names that appear in the Tbox $\mathcal{T}$). One can check that none of the possible substitutions leads to a solution to our matching problem. Hence, the matching problem $Doctor \equiv^?_{\mathcal{T}} Academic$ is unsolvable in the context of a non refreshing semantics. This is not the case, however, if we consider refreshing semantics (i.e., when x and y can be refreshed). Let x_i denotes the i^{th} occurrence of a variable x during an unfolding process. Under refreshing semantics, it is then possible to define a variable substitution σ that alternates the values assignment to x and y as follows :

- $\sigma(x_i) = getPhDIn$ and $\sigma(y_i) = formerly$, for odd i, and
- $\sigma(x_i) = studyIn$ and $\sigma(y_i) = supervisedBy$ for even i.

The resulting concept, $\sigma(Academic) \equiv Person \sqcap \exists getPhDIn.University \sqcap \exists formerly.(\exists studyIn.University \sqcap \exists supervisedBy.(...))$ matches the definition of $Doctor$ (i.e., $Doctor \equiv_{\mathcal{T}} \sigma(Academic)$). Therefore, the matching problem $Doctor \equiv^?_{\mathcal{T}} Academic$ is unsolvable under the non refreshing semantics while it has a solution under the refreshing semantics. Note that, reasoning (e.g., matching or unification) in description logics with refreshing variables is not an easy task. Indeed, as illustrated in our example, it amounts to reasoning on patterns with an infinite number of variables.

To illustrate the second issue, we consider the pattern $Academic2 \equiv Person \sqcap \exists getPhdIn.X \sqcap \exists formerly.(Person \sqcap \exists studyIn.X \sqcap \exists supervisedBy.Academic2)$ and the following matching problem modulo subsumption: $Academic2 \sqsubseteq^?_{\mathcal{T}} FrenchDoctor$. Under the non refreshing semantics, the best matcher is given by the substitution σ defined as: $\sigma(X) = FrenchUniv$ which leads to $\sigma(Academic2) \sqsubseteq FrenchDoctor$ (and more precisely, $\sigma(Academic2) \sqsubset FrenchDoctor$). Under the refreshing semantics, it is possible to find a solution that is more general than $\sigma(Academic2)$. Such a solution is given by a substitution θ defined as follows: $\theta(X_1) = FrenchUniv$ and $\theta(X_i) = University$, for $i > 1$. The resulting concept would achieve equivalence with *FrenchDoctor*. This later solution is interesting since it is closer to the original concept $FrenchDoctor$. Indeed, we have: $\sigma(Academic2) \sqsubset \theta(Academic2) \equiv FrenchDoctor$.

Our goal is to study a new reasoning mechanism, called *weak-subsumption*, in description logics with refreshing variables. Given a description logic $\mathcal{L}$, weak subsumption is informally defined as follows: let $\mathcal{T}$ be an $\mathcal{L}$-TBox and let P, Q be two $\mathcal{L}$-patterns with variables. Then, P is weakly subsumed by Q iff there exists two substitutions ϕ_1 and ϕ_2 s.t $\phi_1(P)$ is subsumed by $\phi_2(Q)$.

Preliminary Results. In [13], we investigated the weak subsumption problem in a restricted framework consisting of the description logic $\mathcal{EL}$ extended with refreshing role variables. More precisely, we introduce a new description logic, called $\mathcal{EL}_{RV}$, that extends the description logic $\mathcal{EL}$ with refreshing role variables.

Our main technical result is to show that testing weak-subsumption between $\mathcal{EL}_{RV}$-patterns is EXPTIME-complete. Our reasoning procedure to test weak subsumption between $\mathcal{EL}_{RV}$ concepts exploits the link between subsumption and the simulation relation between the so-called description graphs introduced in [2] as well as a specific notion of description automata, inspired from fresh variable automata [8], to handle variables with refreshing semantics. The main steps of our approach are:

– We associate with each $\mathcal{EL}_{RV}$-pattern P a description automaton A_P corresponding to a compact representation of all possible instantiations of P.
– We extend the notion of simulation relation, used in [2] to characterize subsumption between $\mathcal{EL}$-patterns. Our main technical result consists in the characterization of weak simulation between $\mathcal{EL}_{RV}$-patterns in terms of existential simulation between $\mathcal{EL}_{RV}$-description graphs.
– We devise an algorithm to test existential simulation between $\mathcal{EL}_{RV}$-description graphs and prove its correctness. We show that the proposed algorithm has exponential time complexity in the worst case and hence is optimal since the weak subsumption problem is EXPTIME-complete.

Future Works. We envision the extension of the approach in three directions: *(i)* extending our framework to handle concept variables, *(ii)* considering additional reasoning mechanisms in this context that go beyond weak-subsumption (e.g., a form of universal or strong subsumption), and *(iii)* considering other description logics such as the logic $\mathcal{FL}_0$ and $\mathcal{ALN}$.

References

1. Baader, F., Küsters, R., Borgida, A., McGuinness, D.: Matching in description logics. Journal of Logic and Computation **9**(3), 411–447 (1999)
2. Baader, F.: Terminological cycles in a description logic with existential restrictions. In: IJCAI. vol. 3, pp. 325–330 (2003)
3. Baader, F., Borgwardt, S., Morawska, B.: Extending unification in $\mathcal{EL}$ towards general tboxes. In: Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning (2012)
4. Baader, F., Gil, O.F., Marantidis, P.: Matching in the description logic $\mathcal{FL}_0$ with respect to general tboxes. In: LPAR. pp. 76–94 (2018)
5. Baader, F., Morawska, B.: Unification in the description logic $\mathcal{EL}$. In: International Conference on Rewriting Techniques and Applications. pp. 350–364. Springer (2009)
6. Baader, F., Morawska, B.: Matching with respect to general concept inclusions in the description logic $\mathcal{EL}$. In: Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz). pp. 135–146. Springer (2014)
7. Baader, F., Narendran, P.: Unification of concept terms in description logics. Journal of Symbolic Computation **31**(3), 277–305 (2001)
8. Belkhir, W., Chevalier, Y., Rusinowitch, M.: Fresh-Variable Automata for Service Composition. In: SYNASC 2013 -15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing. West University of Timisoara Department of Computer Science, IEEE, Timisoara, Romania (Sep 2013), https://hal.inria.fr/hal-00914778, 28 pages. 4 Figures
9. Borgida, A., Küsters, R.: "what's not in a name?" - initial explorations of a structural approach to integrating large concept knowledge-bases (1999)
10. Borgida, A., McGuinness, D.L.: Asking queries about frames. In: Aiello, L.C., Doyle, J., Shapiro, S.C. (eds.) KR'96, Cambridge, Massachusetts, USA, 1996. pp. 340–349. Morgan Kaufmann (1996)
11. Informatik, L., Baader, F., Narendran, P.: Unification of concept terms in description logics (03 1999)
12. Mcguinness, D.L., Borgida, A.: Explaining Reasoning in Description Logics. Ph.D. thesis, USA (1996)
13. T. Ducros, M. Bouet, F.T.: Weak subsumption in the $\mathcal{EL}$-description logic with refreshing variables - extended version. Tech. rep., Technical report of LIMOS, https://hal.archives-ouvertes.fr/hal-03260408 (2021)