Description Logics over Multisets

Yuncheng Jiang[1,2]

[1] School of Computer Science, South China Normal University, Guangzhou 510631, P.R. China
[2] State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing 100190, P.R. China
ycjiang@scnu.edu.cn

Abstract. Description Logics (DLs) are a family of knowledge representation languages that have gained considerable attention the last 20 years. It is well-known that the interpretation domain of classical DLs is a classical set. However, in Science and in the ordinary life the situation is not at all like this. In order to handle these types of knowledge in DLs, in this paper we present a DL framework based on multiset theory. Concretely, we present the DL over multisets $\mathcal{ALC}_{msets}$ which is a semantic extension of the classical DL $\mathcal{ALC}$. The syntax and semantics of $\mathcal{ALC}_{msets}$ are presented. Moreover, we investigate the logical properties of $\mathcal{ALC}_{msets}$ and provide a sound and terminating reasoning algorithm for satisfiability problem of $\mathcal{ALC}_{msets}$.

Keywords: Classical Description Logics; Extended Description Logics; Multisets; Satisfiability

1 Introduction

In the last 20 years a substantial amount of work has been carried out in the context of Description Logics (DLs for short) [1][20]. DLs are a family of logic-based knowledge representation formalisms that are tailored towards representing the terminological knowledge of an application domain in a structured and formally well-understood way. DLs have been applied to numerous problems in computer science such as information integration, databases, software engineering and soft sets. Recent interest in DLs has been spurred by their application in the Semantic Web [2]: the DL $\mathcal{SHOIN}$(**D**) provides the logical underpinning for the Web Ontology Language (OWL), and the DL $\mathcal{SROIQ}$(**D**) is used in OWL 2 [6][11][15][16]. A main point is that DLs are considered as to be attractive logics in knowledge based applications as they are a good compromise between expressive power and computational complexity.

From the semantics of DLs [1] we know that the interpretation domain of classical DLs is a classical set (Zermelo-Fraenkel set) [12]. That is to say, the interpretation of classical DLs is based on classical set theory from a semantics point of view. It is well-known that classical set theory states that a given element can appear only once in a set, it assumes that all mathematical objects occur without

repetition. However, in Science and in the ordinary life the situation is not at all like this. In the physical world it is observed that there is an enormous repetition [7].

As a matter of fact, in order to process the collections with repetition, multi-set theory (MST for short) has been presented and several operations as the addition, the union and the intersection of multisets have been defined and their properties investigated in several papers [3][27][28]. Intuitively, multisets (sometimes also called bags[13][28]) are set-like structures where an element can appear more than once [3]. Thus, a multiset differs from a set in that each element has a multiplicity, which is a natural number indicating (lossely speaking) how many times it is a member of the multiset [7]. We must note that the word multiset was coined by N. G. de Bruijin [18], but the first person that actually used multisets was Richard Dedekind in his well-known paper "Was sind und was sollen die Zahlen?" ("The nature and meaning of numbers") [4]. This paper was published in 1888 [27]. More concretely, a multiset is a collection of objects in which repetition of elements is significant [9]. We confront a number of situations in life when we have to deal with collections of elements in which duplicates are significant. An example may be cited to prove this point. While handling a collection of employees' ages or details of salary in a company, we need to handle entries bearing repetitions and consequently our interest may be diverted to the distribution of elements. In such situations the classical definition of set proves inadequate for the situation presented [9]. Thus, from a practical point of view multisets are very useful structures as they arise in many areas of mathematics and computer science [8][9][19][22][23][27]. A complete survey of the development of multi-set theory can be found in [3].

Naturally, a problem is raised: how we can interpret the concepts and the roles of DLs using multi-set theory? Furthermore, what are the benefits of doing so? After careful thought, we find that it is feasible to interpret the concepts and the roles of DLs using multi-set theory. Moreover, it is a more accurate interpretation for the concepts and the roles of DLs. For example, when we interpret the concept *commended-students* (students who are commended), we can say that *Zhangsan*, *Lisi* and *Wangwu* are the instances of the concept *commended-students*. More formally, we can say $commended\text{-}students^I$={*Zhangsan*, *Lisi*, *Wangwu*} in classical DLs. However, if we consider more accurate situation, e.g., *Zhangsan* is commended three times, *Lisi* is commended twice, and *Wangwu* is commended once, obviously, the classical interpretation of DLs cannot process this situation. Here we can interpret the concept *commended-students* using multi-set theory. Formally, $commended\text{-}students^{MI}$={*Zhangsan*, *Zhangsan*, *Zhangsan*, *Lisi*, *Lisi*, *Wangwu*}, where {*Zhangsan*, *Zhangsan*, *Zhangsan*, *Lisi*, *Lisi*, *Wangwu*} is a multiset.

In this paper we extend DLs allow to express that interpretation of a concept (resp., a role) is not a subset of classical set (traditional interpretation domain Δ^I) (resp., a subset of $\Delta^I \times \Delta^I$) like in classical DLs, but a subset of multisets (resp., a subset of Cartesian product of multisets). That is, we will extend the interpretation domain of DLs to multisets. More concretely, we will present the DL $\mathcal{ALC}_{msets}$, which is a semantic extension of the DL $\mathcal{ALC}$ [10][14][17][24][26] based on multiset theoretic operations presented in [5][9][13]. Moreover, we will provide a sound and incomplete reasoning algorithm for the satisfiability reasoning problem of the DL $\mathcal{ALC}_{msets}$. It is worth noting that classical set is a special case of multisets [9], hence, the DL $\mathcal{ALC}$ [10]

[14][17][24][26] is a special case of the DL $\mathcal{ALC}_{msets}$ presented in this paper from a semantics point of view.

2 Multisets

The current section provides some background on multisets.

A naive concept of multiset was formalized by Blizard [5]. It has the following properties: (i) a multiset is a collection of elements in which certain elements may occur more than once; (ii) occurrences of a particular element in a multiset are indistinguishable; (iii) each occurrence of an element in a multiset contributes to the cardinality of the multiset; (iv) the number of occurrences of a particular element in a multiset is a (finite) positive integer; (v) the number of distinguishable (distinct) elements in a multiset need not be finite; and (vi) a multiset is completely determined if we know the elements that belong to it and the number of times each element belongs to it [9]. In the following, we introduce the basic definitions and notations of multisets [5][9][13].

A collection of elements containing duplicates is called a multiset. Formally, if X is a set of elements, a multiset M drawn from the set X is represented by a function count M or C_M: $X \rightarrow N$ where N represents the set of non-negative integers. For each $x \in X$, $C_M(x)$ is the characteristic value of x in M and indicates the number of occurrences of the element x in M. A multiset M is a set if $\forall x \in X$, $C_M(x)=0$ or 1.

The word "multiset" often shortened to "mset" abbreviates the term "multiple membership set".

Let M_1 and M_2 be two msets drawn from a set X. M_1 is a sub mset of M_2 ($M_1 \subseteq M_2$) if $\forall x \in X$, $C_{M1}(x) \leq C_{M2}(x)$. M_1 is a proper sub mset of M_2 ($M_1 \subset M_2$) if $C_{M1}(x) \leq C_{M2}(x)$ $\forall x \in X$ and there exists at least one $\forall x \in X$ such that $C_{M1}(x) < C_{M2}(x)$. Two msets M_1 and M_2 are equal ($M_1=M_2$) if $M_1 \subseteq M_2$ and $M_2 \subseteq M_1$. An mset M is empty if $\forall x \in X$, $C_M(x)=0$. The cardinality of an mset M drawn from a set X is Card $M=\Sigma_{x \in X} C_M(x)$. It is also denoted by $|M|$.

The insertion of an element x into an mset M results in a new mset M' denoted by $M'=M \oplus x$ such that $C_{M'}(x)=C_M(x)+1$ and $C_{M'}(y)=C_M(y)$ $\forall y \neq x$. Addition of two msets M_1 and M_2 drawn from a set X results in a new mset $M=M_1 \oplus M_2$ such that $\forall x \in X$, $C_M(x)=C_{M1}(x)+C_{M2}(x)$. The removal of an element x from an mset M results in a new mset M' denoted by $M'=M \ominus x$ such that $C_{M'}(x)= \max\{C_M(x)-1, 0\}$ and $C_{M'}(y)=C_M(y)$ $\forall y \neq x$. Subtraction of two msets M_1 and M_2 drawn from a set X results in a new mset $M=M_1 \ominus M_2$ such that $\forall x \in X$, $C_M(x)=\max\{C_{M1}(x)-C_{M2}(x), 0\}$. The union of two msets M_1 and M_2 drawn from a set X is an mset M denoted by $M=M_1 \cup M_2$ such that $\forall x \in X$, $C_M(x)=\max\{C_{M1}(x), C_{M2}(x)\}$. The intersection of two msets M_1 and M_2 drawn from a set X is an mset M denoted by $M=M_1 \cap M_2$ such that $\forall x \in X$, $C_M(x)=\min\{C_{M1}(x), C_{M2}(x)\}$.

Let M be an mset from X with x appearing n times in M. It is denoted by $x \in^n M$. $M=\{k_1/x_1, k_2/x_2, \ldots, k_n/x_n\}$ where M is an mset with x_1 appearing k_1 times, x_2 appearing k_2 times and so on. $[M]_x$ denotes that the element x belongs to the mset M and $|[M]_x|$ denotes the cardinality of an element x in M. The entry of the form $(m/x, n/y)/k$

denotes that x is repeated m times, y is repeated n times and the pair (x, y) is repeated k times. $C_1(x, y)$ denotes the count of the first co-ordinate in the ordered pair (x, y) and $C_2(x, y)$ denotes the count of the second co-ordinate in the ordered pair (x, y).

The mset space X^n is the set of all msets whose elements are in X such that no element in the mset occurs more than n times. The set X^∞ is the set of all msets over a domain X such that there is no limit to the number of occurrences of an element in an mset. If $X=\{x_1, x_2, \ldots, x_k\}$ then $X^n= \{\{n_1/x_1, n_2/x_2, \ldots, n_n/x_n\} \mid \text{for } i=1, 2, \ldots, k; n_i \in \{0, 1, 2, \ldots, n\}\}$.

Let X be a support set and X^n be the mset space defined over X. Then for any mset $M \in X^n$, the complement M^c of M in X^n is an element of X^n such that $\forall x \in X$, $C_{Mc}(x)=n-C_M(x)$.

Let M_1 and M_2 be two msets drawn from a set X, then the Cartesian product of M_1 and M_2 is defined as $M_1 \times M_2=\{(m/x, n/y)/mn \mid x \in^m M_1, y \in^n M_2\}$. The Cartesian product of three or more nonempty msets can be defined by generalizing the definition of the Cartesian product of two msets. Thus the Cartesian product $M_1 \times M_2 \times \ldots \times M_n$ of the nonempty msets $M_1, M_2, \ldots, M_n$ is the mset of all ordered n-tuples $(m_1, m_2, \ldots, m_n)$ where $m_i \in^{ri} M_i$, $i=1, 2, \ldots, n$ and $(m_1, m_2, \ldots, m_n) \in^p M_1 \times M_2 \times \ldots \times M_n$ with $p=\prod r_i$, where $r_i=C_{Mi}(m_i)$, and $i=1, 2, \ldots, n$.

A sub mset R of $M \times M$ is said to be an mset relation on M if every member $(m/x, n/y)$ of R has a count $C_1(x, y).C_2(x, y)$. We denote m/x related to n/y by $m/xRn/y$.

The domain and range of the mset relation R on M is defined as follows: Dom $R=\{x \in^r M \mid \exists y \in^s M$ such that $r/xRs/y\}$ where $C_{DomR}(x)=\sup\{C_1(x, y) \mid x \in^r M\}$, Ran $R=\{y \in^s M \mid \exists x \in^r M$ such that $r/xRs/y\}$ where $C_{RanR}(y)=\sup\{C_2(x, y) \mid y \in^s M\}$.

3 The $\mathcal{ALC}_{msets}$ DL

In the current section we will present the description logic over multisets $\mathcal{ALC}_{msets}$, which is a semantic extension of the $\mathcal{ALC}$ [1][24]. Concretely, we first define its syntax and semantics. Then, we discuss its logical properties.

3.1 Syntax and Semantics

As usual, we consider an alphabet of distinct concept names (**C**), role names (**R**) and individual names (**I**). The abstract syntax of $\mathcal{ALC}_{msets}$-concepts and $\mathcal{ALC}_{msets}$-roles is the same as that of $\mathcal{ALC}$ [1][24]; however, their semantics is based on interpretations on multisets (msets interpretations for short) (see below). Similarly, $\mathcal{ALC}_{msets}$ keeps the same syntax of terminological axioms (concept inclusions and concept equations) as that of $\mathcal{ALC}$. Interestingly, $\mathcal{ALC}_{msets}$ extends $\mathcal{ALC}$ assertions (concept assertions and role assertions) into mset assertions, where individuals containing duplicates can appear.

In the following, we give the semantics of $\mathcal{ALC}_{msets}$-concepts and $\mathcal{ALC}_{msets}$-roles formally.

An mset interpretation is a pair $MI = (\Delta^{MI}, \bullet^{MI})$, where Δ^{MI} is a non-empty mset (the interpretation domain), and $\bullet^{MI}$ is an interpretation function that assigns each

atomic concept (concept name) $A \in \mathbf{C}$ to a set $A^{MI} \subseteq \Delta^{MI}$, each atomic role (role name) $R \in \mathbf{R}$ (note that in $\mathcal{ALC}_{msets}$ roles are always atomic) to a binary relation $R^{MI} \subseteq \Delta^{MI} \times \Delta^{MI}$, and each individual name $m/a \in \mathbf{I}$ to an element $a^{MI} \in {}^{m}\Delta^{MI}$. This interpretation function is extended to $\mathcal{ALC}_{msets}$ concept descriptions as follows:

- $\top^{MI} = \Delta^{MI}$;
- $\bot^{MI} = \phi$;
- $(\neg C)^{MI} = \Delta^{MI} \Theta C^{MI}$;
- $(C \sqcap D)^{MI} = C^{MI} \cap D^{MI}$;
- $(C \sqcup D)^{MI} = C^{MI} \cup D^{MI}$;
- $(\exists R.C)^{MI} = \{a \in {}^{m}\Delta^{MI} \mid \exists b \in {}^{n}\Delta^{MI}, (m/a, n/b) \in {}^{mn}R^{MI} \wedge b \in {}^{n}C^{MI}\}$;
- $(\forall R.C)^{MI} = \{a \in {}^{m}\Delta^{MI} \mid \forall b \in {}^{n}\Delta^{MI}, (m/a, n/b) \in {}^{mn}R^{MI} \rightarrow b \in {}^{n}C^{MI}\}$.

Note that in this paper we restrict the interpretation domain to be finite. This is not a severe limitation as it is hard to imagine an application involving infinite interpretation domains.

An $\mathcal{ALC}_{msets}$ knowledge base $\mathcal{KB}$ is also composed of a TBox and an ABox. A TBox $\mathcal{T}$ is a finite, possibly empty, set of terminological axioms that could be a combination of concept inclusions of the form $\langle C \sqsubseteq D \rangle$ and concept equations of the form $\langle C \equiv D \rangle$, where C and D are concept descriptions. An mset interpretation MI satisfies $\langle C \sqsubseteq D \rangle$ if $C^{MI} \subseteq D^{MI}$, and it satisfies $\langle C \equiv D \rangle$ if $C^{MI} = D^{MI}$ (i.e., $C^{MI} \subseteq D^{MI}$ and $D^{MI} \subseteq C^{MI}$). An mset interpretation MI satisfies a TBox $\mathcal{T}$ iff MI satisfies every axiom in $\mathcal{T}$; in this case, we say that MI is a model of $\mathcal{T}$.

An ABox $\mathcal{A}$ includes of a set of mset assertions that could be a combination of concept assertions of the form $\langle m/a : C \rangle$ and role assertions of the form $\langle (m/a, n/b) : R \rangle$, where a and b are individuals, C is a concept, and R is a role. An mset interpretation MI satisfies $\langle m/a : C \rangle$ if $a^{MI} \in {}^{m}C^{MI}$, and it satisfies $\langle (m/a, n/b) : R \rangle$ if $(m/a^{MI}, n/b^{MI}) \in {}^{mn}R^{MI}$. An mset interpretation MI satisfies an ABox $\mathcal{A}$ iff MI satisfies every mset assertion in $\mathcal{A}$ w.r.t. a TBox $\mathcal{T}$; in this case, we say that MI is a model of $\mathcal{A}$ w.r.t. $\mathcal{T}$.

An mset interpretation MI satisfies (or is a model of) a knowledge base $\mathcal{KB} = \langle \mathcal{T}, \mathcal{A} \rangle$ (denoted $MI \models \mathcal{KB}$), iff it satisfies both components of $\mathcal{KB}$; in this case, we say that MI is a model of $\mathcal{KB}$. The knowledge base $\mathcal{KB}$ is consistent if there exists an mset interpretation MI that satisfies $\mathcal{KB}$. We say $\mathcal{KB}$ is inconsistent otherwise.

Description logics over multisets should provide their users with reasoning capabilities that allow them to derive implicit knowledge from the one explicitly represented. In the following we will define the most important reasoning problems of the $\mathcal{ALC}_{msets}$ DL.

Let $\mathcal{T}$ be a TBox, $\mathcal{A}$ an ABox, C, D concept descriptions, and a an individual name. The definitions of the main reasoning problems of the $\mathcal{ALC}_{msets}$ DL are as follows:

- C is subsumed by D w.r.t. $\mathcal{T}$ ($\langle C \sqsubseteq_{\mathcal{T}} D \rangle$) iff $C^{MI} \subseteq D^{MI}$ for all models MI of $\mathcal{T}$;
- C is equivalent to D w.r.t. $\mathcal{T}$ ($\langle C \equiv_{\mathcal{T}} D \rangle$) iff $C^{MI} = D^{MI}$ for all models MI of $\mathcal{T}$;
- C is satisfiable w.r.t. $\mathcal{T}$ iff $C^{MI} \neq \phi$ for some model MI of $\mathcal{T}$;
- $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$ iff it has a model that is also a model of $\mathcal{T}$;

- m/a is an instance of C w.r.t. $\mathcal{A}$ and $\mathcal{T}$ ($\mathcal{A}\models_{\mathcal{T}}\langle m/a{:}C\rangle$) iff $a^{MI}\in^{m}C^{MI}$ for all models MI of $\mathcal{T}$ and $\mathcal{A}$.

One might think that, in order to realize the reasoning component of an $\mathcal{ALC}_{msets}$ system, one need to design and implement five algorithms, each solving one of the above reasoning problems. Fortunately, this is not the case since there exist some polynomial time reductions (see Section 3.2).

3.2 Logical Properties

It can be easily shown that $\mathcal{ALC}_{msets}$ is a sound extension of $\mathcal{ALC}$, in the sense that the mset interpretations coincide with the traditional interpretations if we restrict the interpretation domain Δ^{MI} to a classical set. However, since $\mathcal{ALC}_{msets}$ is based on multiset theory, some properties which are different from $\mathcal{ALC}$ are obtained. Of course, some properties are the same as that of $\mathcal{ALC}$. In the following, we will discuss these properties.

The first ones are straightforward: $\langle\neg\top\equiv\bot\rangle$, $\langle\neg\bot\equiv\top\rangle$, $\langle C\sqcap\top\equiv C\rangle$, $\langle C\sqcup\bot\equiv C\rangle$, $\langle C\sqcap\bot\equiv\bot\rangle$, $\langle C\sqcup\top\equiv\top\rangle$ and $\langle\exists R.\bot\equiv\bot\rangle$, where C is a concept, R is a role.

The following properties show that some interesting equivalences hold in $\mathcal{ALC}_{msets}$.

Proposition 1. Let C, C_1, C_2, C_3 and D be five concepts. Then

(1) $\langle\neg\neg C\equiv C\rangle$, $\langle C\sqcap C\equiv C\rangle$, $\langle C\sqcup C\equiv C\rangle$;
(2) $\langle\neg(C\sqcap D)\equiv\neg C\sqcup\neg D\rangle$, $\langle\neg(C\sqcup D)\equiv\neg C\sqcap\neg D\rangle$;
(3) $\langle C_1\sqcup(C_2\sqcap C_3)\equiv(C_1\sqcup C_2)\sqcap(C_1\sqcup C_3)\rangle$, $\langle C_1\sqcap(C_2\sqcup C_3)\equiv(C_1\sqcap C_2)\sqcup(C_1\sqcap C_3)\rangle$.

Note 1. Please note that the following properties are satisfied in $\mathcal{ALC}$, however, these properties are not satisfied in $\mathcal{ALC}_{msets}$:

$\langle(C\sqcap\neg C)\equiv\bot\rangle$, $\langle(C\sqcup\neg C)\equiv\top\rangle$, $\langle\forall R.\top\equiv\top\rangle$, $\langle\neg(\forall R.C)\equiv\exists R.\neg C\rangle$, $\langle\neg(\exists R.C)\equiv\forall R.\neg C\rangle$, $\langle(\exists R.C)\sqcup(\exists R.D)\equiv\exists R.(C\sqcup D)\rangle$, and $\langle(\forall R.C)\sqcap(\forall R.D)\equiv\forall R.(C\sqcap D)\rangle$.

There are two interesting remarks here. Firstly, in $\mathcal{ALC}$, we can assume concepts to be in negation normal form (NNF), i.e., negation signs occur immediately in front of concept names only. However, in $\mathcal{ALC}_{msets}$, we cannot do this translation due to $\langle\neg(\forall R.C)\not\equiv\exists R.\neg C\rangle$ and $\langle\neg(\exists R.C)\not\equiv\forall R.\neg C\rangle$. Secondly, in $\mathcal{ALC}$, an ABox $\mathcal{A}$ contains a clash iff $\{A(a), \neg A(a)\}\subseteq\mathcal{A}$ for some individual name a and some concept name A. However, in $\mathcal{ALC}_{msets}$, we cannot use this definition due to $\langle(C\sqcap\neg C)\not\equiv\bot\rangle$ and $\langle(C\sqcup\neg C)\not\equiv\top\rangle$. For example, let $\Delta^{MI}=\{6/a, 8/b\}$ and $\{\langle 3/a{:}C\rangle, \langle 4/b{:}C\rangle\}\subseteq\mathcal{A}$. Since $\langle 3/a{:}C\rangle$ and $\langle 4/b{:}C\rangle$, then we have $\langle 3/a{:}\neg C\rangle, \langle 4/b{:}\neg C\rangle\in\mathcal{A}$. That is, $\{\langle 3/a{:}C\rangle, \langle 3/a{:}\neg C\rangle, \langle 4/b{:}C\rangle, \langle 4/b{:}\neg C\rangle\}\subseteq\mathcal{A}$.

The properties of the polynomial time reductions for reasoning problems are as follows.

Proposition 2. Let $\mathcal{T}$ be a TBox, $\mathcal{A}$ an ABox, C, D concept descriptions, and a an individual name. Then

(1) $\langle C \sqsubseteq_{\mathcal{T}} D \rangle$ iff $\langle C \sqcap D \equiv_{\mathcal{T}} C \rangle$;

(2) $\langle C \equiv_{\mathcal{T}} D \rangle$ iff $\langle C \sqsubseteq_{\mathcal{T}} D \rangle$ and $\langle D \sqsubseteq_{\mathcal{T}} C \rangle$;

(3) C is satisfiable w.r.t. $\mathcal{T}$ iff not $\langle C \sqsubseteq_{\mathcal{T}} \bot \rangle$;

(4) C is satisfiable w.r.t. $\mathcal{T}$ iff there exist $m>0$ and individual a such that $\{\langle m/a:C \rangle\}$ is consistent w.r.t. $\mathcal{T}$;

(5) $\mathcal{A}$ is consistent w.r.t. $\mathcal{T}$ iff $\mathcal{A} \not\models_{\mathcal{T}} \langle m/a:\bot \rangle$ for any $m>0$ and individual a.

Note 2. It needs to be noted that the polynomial time reductions for instance problem to (in)consistency (i.e., $\mathcal{A} \models_{\mathcal{T}} \langle m/a:C \rangle$ iff $\mathcal{A} \cup \{\langle m/a:\neg C \rangle\}$ is inconsistent w.r.t. $\mathcal{T}$) and subsumption problem to (un)satisfiability (i.e., $\langle C \sqsubseteq_{\mathcal{T}} D \rangle$ iff $C \sqcap \neg D$ is unsatisfiable w.r.t. $\mathcal{T}$), are satisfied in $\mathcal{ALC}$, however, these reductions are not satisfied in $\mathcal{ALC}_{msets}$.

Lastly, we have to point out that in the rest of this paper we only consider unfoldable TBoxes. More concretely, a concept definition is of the form $\langle A \equiv C \rangle$ where A is a concept name and C is a concept description. Given a set $\mathcal{T}$ of concept definitions, we say that the concept name A directly uses the concept name B if $\mathcal{T}$ contains a concept definition $\langle A \equiv C \rangle$ such that B occurs in C. Let uses be the transitive closure of the relation "directly uses". We say that $\mathcal{T}$ is cyclic if there is a concept name A that uses itself, and acyclic otherwise. A TBox $\mathcal{T}$ is a finite, possibly empty, set of terminological axioms of the form $\langle A \sqsubseteq C \rangle$, called inclusion introductions, and of the form $\langle A \equiv C \rangle$, called equivalence introductions. A TBox is unfoldable if it contains no cycles and contains only unique introductions, i.e., terminological axioms with only concept names appearing on the left hand side and, for each concept name A, there is at most one axiom in $\mathcal{T}$ of which A appears on the left side.

In classical DLs [1], a knowledge base with an unfoldable TBox can be transformed into an equivalent one with an empty TBox by a transformation called unfolding or expansion [21][25]: Concept inclusion introductions $\langle A \sqsubseteq C \rangle$ are replaced by concept equivalence introductions $\langle A \equiv A' \sqcap C \rangle$, where A' is a new concept name, which stands for the qualities that distinguish the elements of A from the other elements of C. Subsequently, if C is a complex concept expression, which is defined in terms of concept names, defined in the TBox, we replace their definitions in C. It has been proved that the initial TBox with the expanded one are equivalent.

In DLs over msets such as $\mathcal{ALC}_{msets}$ presented in this paper, we also can prove that a knowledge base with an unfoldable TBox can be transformed into an equivalent one with an empty TBox.

Firstly, we can transform an $\mathcal{ALC}_{msets}$-TBox $\mathcal{T}$ into a regular $\mathcal{ALC}_{msets}$-TBox $\mathcal{T}'$, containing equivalence introductions only, such that $\mathcal{T}'$ is equivalent to $\mathcal{T}$ in a sense that will be specified below. We obtain $\mathcal{T}'$ from $\mathcal{T}$ by choosing for every concept inclusion introduction $\langle A \sqsubseteq C \rangle$ in $\mathcal{T}$ a new concept name A' and by replacing the inclusion introduction $\langle A \sqsubseteq C \rangle$ with the equivalence introduction $\langle A \equiv A' \sqcap C \rangle$. The TBox $\mathcal{T}'$ is the normalization of $\mathcal{T}$.

Now we show that $\mathcal{T}$ and $\mathcal{T}'$ are equivalent.

Proposition 3. Let $\mathcal{T}$ be a TBox and $\mathcal{T}'$ its normalization. Then

(1) Every model of $\mathcal{T}'$ is a model of $\mathcal{T}$.

(2) For every model MI of $\mathcal{T}$ there is a model MI' of $\mathcal{T}'$ that has the same domain as MI and agrees with MI on the concept names and roles in $\mathcal{T}$.

Thus, in theory, inclusion introductions do not add to the expressivity of TBoxes. However, in practice, they are a convenient means to introduce concepts into a TBox that cannot be defined completely. In fact, this case is the same as classical DLs [1].

Now we show that, if $\mathcal{T}$ is an unfoldable TBox, we can always reduce reasoning problems w.r.t. $\mathcal{T}$ to problems w.r.t. the empty TBox. Instead of saying "w.r.t. ϕ" one usually says "without a TBox", and omits the index $\mathcal{T}$ for subsumption, equivalence, and instance, i.e., writes $\equiv$, $\sqsubseteq$, $\models$ instead of $\equiv_{\mathcal{T}}$, $\sqsubseteq_{\mathcal{T}}$, and $\models_{\mathcal{T}}$. As we have seen in Proposition 3, $\mathcal{T}$ is equivalent to its expansion $\mathcal{T}'$. Recall that in the expansion every equivalence introduction $\langle A \equiv D \rangle$ such that D contains only concept names, but no concept descriptions. Now, for each concept description C we define the expansion of C w.r.t. $\mathcal{T}$ as the concept description C'' that is obtained from C by replacing each occurrence of a concept name A in C by the concept description D, where $\langle A \equiv D \rangle$ is the equivalence introduction of A in $\mathcal{T}'$, the expansion of $\mathcal{T}$.

Proposition 4. Let $\mathcal{T}$ be an unfoldable TBox, C, D concept descriptions, C'' expansion of C, and D'' expansion of D. Then

(1) $\langle C \equiv_{\mathcal{T}} C'' \rangle$;
(2) C is satisfiable w.r.t. $\mathcal{T}$ iff C'' is satisfiable;
(3) $\langle C \sqsubseteq_{\mathcal{T}} D \rangle$ iff $\langle C'' \sqsubseteq D'' \rangle$;
(4) $\langle C \equiv_{\mathcal{T}} D \rangle$ iff $\langle C'' \equiv D'' \rangle$.

4 Reasoning in $\mathcal{ALC}_{msets}$

In this section, we will provide a detailed presentation of the reasoning algorithm for the $\mathcal{ALC}_{msets}$-satisfiability problem and the properties for the termination and soundness of the procedure. There is one point we have to point out here. Since we restrict the maximal number of occurrences of an element (i.e., an individual) in a multiset (i.e., subset of interpretation domain), it is obvious to know that the satisfiability reasoning algorithm (see below) is incomplete.

In the following, we will present a tableau algorithm for testing satisfiability of an $\mathcal{ALC}_{msets}$-concept. Before we can describe the tableau-based satisfiability algorithm for $\mathcal{ALC}_{msets}$ more formally, we need to introduce some basic notions firstly.

A constraint (denoted by α) is an expression of the form $\langle m/a{:}C \rangle$, or $\langle (m/a, n/b){:}R \rangle$, where a and b are individuals, C is a concept, and R is a role. Our calculus, determining whether a finite set $\mathcal{S}$ of constraints or not, is based on a set of constraint propagation rules transforming a set $\mathcal{S}$ of constraints into "simpler" satisfiability preserving sets $\mathcal{S}_i$ until either all $\mathcal{S}_i$ contain a clash (indicating that from all the $\mathcal{S}_i$ no model of $\mathcal{S}$ can be build) or some $\mathcal{S}_i$ is completed and clash-free, that is, no rule can

further be applied to $\mathcal{S}_i$ and $\mathcal{S}_i$ contains no clash (indicating that from $\mathcal{S}_i$ a model of $\mathcal{S}$ can be build). A set of constraints $\mathcal{S}$ contains a clash iff $\{\langle m/a{:}C\rangle, \langle 0/a{:}C\rangle\} \subseteq \mathcal{S}$ for some $m>0$, individual a, and concept description C.

The $\rightarrow_{\neg}$-rule

Condition: $\mathcal{S}_i$ contains $\langle m/a{:}\neg C\rangle$, but it does not contain $\langle 1/a{:}C\rangle$, $\langle 2/a{:}C\rangle$, …, or $\langle nmax-m/a{:}C\rangle$.

Action: $\mathcal{S}_{i,1}=\mathcal{S}_i\cup\{\langle 1/a{:}C\rangle\}$, $\mathcal{S}_{i,2}=\mathcal{S}_i\cup\{\langle 2/a{:}C\rangle\}$, …, $\mathcal{S}_{i,nmax-m}=\mathcal{S}_i\cup\{\langle nmax-m/a{:}C\rangle\}$.

The $\rightarrow_{\sqcap}$-rule

Condition: $\mathcal{S}_i$ contains $\langle m/a{:}C_1\sqcap C_2\rangle$, but neither $\{\langle m/a{:}C_1\rangle, \langle j/a{:}C_2\rangle\}$ nor $\{\langle m/a{:}C_2\rangle, \langle j/a{:}C_1\rangle\}$, where $m\leq j\leq nmax$.

Action: $\mathcal{S}_{i,1}'=\mathcal{S}_i\cup\{\langle m/a{:}C_1\rangle, \langle m/a{:}C_2\rangle\}$, $\mathcal{S}_{i,2}'=\mathcal{S}_i\cup\{\langle m/a{:}C_1\rangle, \langle m+1/a{:}C_2\rangle\}$, ..., $\mathcal{S}_{i,nmax+1}'=\mathcal{S}_i\cup\{\langle m/a{:}C_1\rangle, \langle nmax/a{:}C_2\rangle\}$, $\mathcal{S}_{i,1}''=\mathcal{S}_i\cup\{\langle m/a{:}C_2\rangle, \langle m/a{:}C_1\rangle\}$, $\mathcal{S}_{i,2}''=\mathcal{S}_i\cup\{\langle m/a{:}C_2\rangle, \langle m+1/a{:}C_1\rangle\}$, ..., $\mathcal{S}_{i,nmax+1}''=\mathcal{S}_i\cup\{\langle m/a{:}C_2\rangle, \langle nmax/a{:}C_1\rangle\}$.

The $\rightarrow_{\sqcup}$-rule

Condition: $\mathcal{S}_i$ contains $\langle m/a{:}C_1\sqcup C_2\rangle$, but neither $\{\langle m/a{:}C_1\rangle, \langle j/a{:}C_2\rangle\}$ nor $\{\langle m/a{:}C_2\rangle, \langle j/a{:}C_1\rangle\}$, where $1\leq j\leq m$.

Action: $\mathcal{S}_{i,1}'=\mathcal{S}_i\cup\{\langle m/a{:}C_1\rangle, \langle m/a{:}C_2\rangle\}$, $\mathcal{S}_{i,2}'=\mathcal{S}_i\cup\{\langle m/a{:}C_1\rangle, \langle m-1/a{:}C_2\rangle\}$, ..., $\mathcal{S}_{i,m}'=\mathcal{S}_i\cup\{\langle m/a{:}C_1\rangle, \langle 1/a{:}C_2\rangle\}$, $\mathcal{S}_{i,1}''=\mathcal{S}_i\cup\{\langle m/a{:}C_2\rangle, \langle m/a{:}C_1\rangle\}$, $\mathcal{S}_{i,2}''=\mathcal{S}_i\cup\{\langle m/a{:}C_2\rangle, \langle m-1/a{:}C_1\rangle\}$, ..., $\mathcal{S}_{i,m}''=\mathcal{S}_i\cup\{\langle m/a{:}C_2\rangle, \langle 1/a{:}C_1\rangle\}$.

The $\rightarrow_{\exists}$-rule

Condition: $\mathcal{S}_i$ contains $\langle m/a{:}\exists R.C\rangle$, but there are no individuals $1/b$, $2/b$, …, $nmax/b$ such that $\langle 1/b{:}C\rangle$ and $\langle (m/a,\ 1/b){:}R\rangle$, $\langle 2/b{:}C\rangle$ and $\langle (m/a,\ 2/b){:}R\rangle$, …, or $\langle nmax/b{:}C\rangle$ and $\langle (m/a,\ nmax/b){:}R\rangle$ are in $\mathcal{S}_i$.

Action: $\mathcal{S}_{i,1}=\mathcal{S}_i\cup\{\langle 1/b{:}C\rangle, \langle (m/a,\ 1/b){:}R\rangle\}$, $\mathcal{S}_{i,2}=\mathcal{S}_i\cup\{\langle 2/b{:}C\rangle, \langle (m/a,\ 2/b){:}R\rangle\}$, …, $\mathcal{S}_{i,nmax}=\mathcal{S}_i\cup\{\langle nmax/b{:}C\rangle, \langle (m/a,\ nmax/b){:}R\rangle\}$, where $1/b$, $2/b$, ..., $nmax/b$ are individuals not occurring in $\mathcal{S}_i$.

The $\rightarrow_{\forall}$-rule

Condition: $\mathcal{S}_i$ contains $\langle m/a{:}\forall R.C\rangle$ and $\langle (m/a,\ n/b){:}R\rangle$, but it does not contain $\langle n/b{:}C\rangle$.

Action: $\mathcal{S}_i'=\mathcal{S}_i\cup\{\langle n/b{:}C\rangle\}$.

Fig. 1. Transformation rules of the satisfiability algorithm

The tableau-based satisfiability algorithm for $\mathcal{ALC}_{msets}$ works as follows. Let C by an $\mathcal{ALC}_{msets}$-concept. In order to test satisfiability of C, the algorithm starts with a finite set of constraints $\{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_{\text{nmax}}\}$, and applies satisfiability preserving transformation rules (see Figure 1) (in arbitrary order) to the set of constraints $\mathcal{S}_i$ ($1\leq i\leq nmax$) until no more rules apply, where $\mathcal{S}_1=\{\langle 1/a{:}C\rangle\}$, $\mathcal{S}_2=\{\langle 2/a{:}C\rangle\}$, …, $\mathcal{S}_{nmax}=\{nmax/a{:}C\}$. If the "complete" constraint obtained this way does not contain an

obvious contradiction (called clash), then $\mathcal{S}$ is consistent (and thus C is satisfiable), and inconsistent (unsatisfiable) otherwise. The transformation rules that handle negation, conjunction, disjunction, and exists restrictions are non-deterministic in the sense that a given set of constraints is transformed into finitely many new sets of constraints such that the original set of constraints is consistent iff one of the new sets of constraints is so. For this reason we will consider finite sets of constraints $\mathcal{S}=\{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_k\}$ instead of the original set of constraints $\{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_{nmax}\}$, where $k \geq nmax$. Such a set is consistent iff there is some i, $1 \leq i \leq k$, such that $\mathcal{S}_i$ is consistent. A rule of Figure 1 is applied to a given finite set of constraints $\mathcal{S}$ as follows: it takes an element $\mathcal{S}_i$ of $\mathcal{S}$, and replaces it by one set of constraints $\mathcal{S}_i'$, by two sets of constraints $\mathcal{S}_i'$ and $\mathcal{S}_i''$, or by finitely many sets of constraints $\mathcal{S}_{i,j}$.

Termination and soundness of the procedures can be shown.

Proposition 5 (Termination)**.** Let C be an $\mathcal{ALC}_{msets}$-concept. There cannot be some infinite sequences of rule applications

$\{\langle j/a:C\rangle\} \rightarrow \mathcal{S}_1 \rightarrow \mathcal{S}_2 \rightarrow \ldots$, where $1 \leq j \leq nmax$.

Proof. The main reasons for this proposition to hold are the following.

(1) The original sets of constraints $\{\langle j/a:C\rangle\}$ is finite. Namely, there exist $nmax$ original sets of constraints $\{\langle 1/a:C\rangle\}$, $\{\langle 2/a:C\rangle\}$, …, $\{\langle nmax/a:C\rangle\}$.

(2) Without loss of generality, we consider the original set of constraints $\{\langle j/a:C\rangle\}$. Let $\mathcal{S}'$ be a set of constraints contained in $\mathcal{S}_i$ for some $i \geq 1$. For every individual $m/b \neq j/a$ occurring in $\mathcal{S}'$, there is a unique sequence $R_1, \ldots, R_k$ ($k \geq 1$) of role names and a unique sequence of individuals of the form $1/b_1$, $1/b_2$, …, $1/b_{k-1}$, or $1/b_1$, $1/b_2$, …, $2/b_{k-1}$, …, or $1/b_1$, $1/b_2$, …, $nmax/b_{k-1}$, …, or $nmax/b_1$, $nmax/b_2$, …, $nmax/b_{k-1}$, such that $\{\langle (j/a, 1/b_1):R_1\rangle, \langle (1/b_1, 1/b_2):R_2\rangle, \ldots, \langle (1/b_{k-1}, m/b):R_k\rangle\} \subseteq \mathcal{S}'$, $\{\langle (j/a, 1/b_1):R_1\rangle, \langle (1/b_1, 1/b_2):R_2\rangle, \ldots, \langle (2/b_{k-1}, m/b):R_k\rangle\} \subseteq \mathcal{S}'$, …, or $\{\langle (j/a, nmax/b_1):R_1\rangle, \langle (nmax/b_1, nmax/b_2):R_2\rangle, \ldots, \langle (nmax/b_{k-1}, m/b):R_k\rangle\} \subseteq \mathcal{S}'$. In this case, we say that m/b occurs on the level k in $\mathcal{S}'$.

(3) If $\langle m/b:C'\rangle \in \mathcal{S}'$ for an individual m/b on level k, then the maximal role depth of C' (i.e., the maximal nesting of constructors involving roles) is bounded by the maximal role depth of C minus k. Consequently, the level of any individual in $\mathcal{S}'$ is bounded by the maximal role depth of C.

(4) If $\langle m/b:C'\rangle \in \mathcal{S}'$, then C' is a subdescription of C. Consequently, the number of different concept assertions on m/b is bounded by the size of C.

(5) The number of different role successors of m/b in $\mathcal{S}'$ (i.e., individuals l/c such that $\langle (m/b, l/c):R\rangle \in \mathcal{S}'$ for a role name R) is bounded by the number of different existential restrictions in C.

Proposition 6 (Soundness)**.** Assume that $\mathcal{S}'$ is obtained from the finite set of constraints $\mathcal{S}$ by application of a transformation rule. If $\mathcal{S}$ is consistent, then $\mathcal{S}'$ is consistent.

Proof. [Sketch] Given the termination property (see Proposition 5), it is easily verified, by case analysis, that the transformation rules of the satisfiability algorithm are sound. For example, the $\rightarrow_{\neg}$-rule: Assume that MI is an mset interpretation satisfying $\langle m/a{:}\neg C\rangle$, where $0<m\leq nmax$. Let us show that MI satisfies $\langle 1/a{:}C\rangle$, $\langle 2/a{:}C\rangle$, …, or $\langle nmax-m/a{:}C\rangle$. Since MI satisfies $\langle m/a{:}\neg C\rangle$, by the semantics of $\neg C$ we have that $a\in^{m}(\neg C)^{MI}=\Delta^{MI}\Theta C^{MI}$. Since the maximal number of occurrences of a in Δ^{MI} is $nmax$, thus, by the definition of subtraction of two msets, we know that the number of occurrences of a in C^{MI} is 1, 2, …, or $nmax-m$. That is, $a\in^{1}C^{MI}$, $a\in^{2}C^{MI}$, …, or $a\in^{nmax-m}C^{MI}$. Therefore, MI satisfies $\langle 1/a{:}C\rangle$, $\langle 2/a{:}C\rangle$, …, or $\langle nmax-m/a{:}C\rangle$.

5 Conclusion

We present a DL framework based on multiset theory. Our main feature is that we extend classical DLs allow to express that interpretation of a concept (resp., a role) is not a subset of classical set (resp., a subset of Cartesian product of sets) like in classical DLs, but a subset of multisets (resp., a subset of Cartesian product of multisets). To the best of our knowledge, this is the first attempt in this direction.

Current research effort is to implement the reasoning algorithm and to perform an empirical evaluation in real scenarios. An interesting topic of future research is to study the complexity and optimization techniques of reasoning in DLs over multisets such as $\mathcal{ALC}_{msets}$. Furthermore, additional research effort can be focused on the reasoning algorithms for the (very) expressive DLs over multisets such as $\mathcal{SROIQ}_{msets}$.

Acknowledgments. The works described in this paper are supported by The National Natural Science Foundation of China under Grant No. 60663001; The Foundation of the State Key Laboratory of Computer Science of Chinese Academy of Sciences under Grant No. SYSKF0904; The Natural Science Foundation of Guangdong Province of China under Grant No. 10151063101000031; The Natural Science Foundation of Guangxi Province of China under Grant No. 0991100.

References

1. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.,: The Description Logic Handbook: Theory, Implementation and Applications, 2nd Edition. Cambridge University Press, Cambridge (2007)
2. Berners-Lee, T., Hendler, J., Lassila, O.,: The Semantic Web. Scientific American 284, 34--43 (2001)
3. Blizard, W.D.,: The Development of Multiset Theory. Modern Logic 1, 319--352 (1991)
4. Blizard, W.D.,: Dedeking Multisets and Functions Shells. Theoretical Computer Science, 110, 79--98 (1993)
5. Blizard, W.D.,: Multiset Theory. Notre Dame Journal of Logic 30, 36--65 (1989)
6. Bobillo, F., Delgado, M., Gomez-Romero, J., Straccia, U.,: Fuzzy Description Logics under Gödel Semantics. International Journal of Approximate Reasoning 50, 494--514 (2009)
7. Casasnovas, J., Mayor, G.,: Discrete t-norms and Operations on Extended Multisets. Fuzzy Sets and Systems 159, 1165--1177 (2008)

8. Csuhaj-Varju, E., Martin-Vide, C., Mitrana, V.,: Multiset Automata. In: Calude, C.S., Paun, G., Rozenberg, G., Salomaa, A. (eds.). Multiset Processing: Mathematical, Computer Science, and Molecular Computing Points of View. LNCS, vol. 2235, pp. 69--83. Springer, Heidelberg (2001)
9. Girish, K.P., John, S.J.,: Relations and Functions in Multiset Context. Information Sciences 179, 758--768 (2009)
10. Haarslev, V., Pai, H.I., Shiri, N.,: A Formal Framework for Description Logics with Uncertainty. International Journal of Approximate Reasoning 50, 1399--1415 (2009)
11. Horrocks, I., Kutz, O., Sattler, U.,: The even more irresistible SROIQ. In: Proceedings of the 10th International Conference of Knowledge Representation and Reasoning, pp. 57--67. AAAI Press (2006)
12. Hrbacek, K., Jech, T.,: Introduction to Set Theory. Marcel Dekker Inc. (1984)
13. Jena, S.P., Ghosh, S.K., Tripathy, B.K.,: On the Theory of Bags and Lists. Information Sciences 132, 241--254 (2001)
14. Jiang, Y., Tang, Y., Wang, J., Tang, S.,: Reasoning within Intuitionistic Fuzzy Rough Description Logics. Information Sciences 179, 2362--2378 (2009)
15. Jiang, Y., Wang, J., Deng, P., Tang, S.,: Reasoning within Expressive Fuzzy Rough Description Logics. Fuzzy Sets and Systems 160, 3403--3424 (2009)
16. Jiang, Y., Tang, Y., Wang, J., Deng, P., Tang, S.,: Expressive Fuzzy Description Logics over Lattices. Knowledge-Based Systems 23, 150--161 (2010)
17. Jiang, Y., Wang, J., Tang, S., Xiao, B.,: Reasoning with Rough Description Logics: An Approximate Concepts Approach. Information Sciences 179, 600--612 (2009)
18. Knuth, D.E.,: The Art of Computer Programming, vol. 2: Seminumerical Algorithms. Addison-Wesley (1981)
19. Lamperti, G., Melchiori, M., Zanella, M.,: On Multisets in Database Systems. In: Calude, C.S., Paun, G., Rozenberg, G., Salomaa A. (eds.). Multiset Processing: Mathematical, Computer Science, and Molecular Computing Points of View. LNCS, vol. 2235, pp. 147--215. Springer, Heidelberg (2001)
20. Lutz, C.,: NEXP TIME-complete Description Logics with Concrete Domains. ACM Transactions on Computational Logic 5, 669--705 (2004)
21. Nebel, B.,: Terminological Reasoning is Inherently Intractable. Artificial Intelligence 43, 235--249 (1990)
22. Miyamoto, S.,: Information Clustering Based on Fuzzy Multisets. Information Processing and Management 39, 195--213 (2003)
23. Miyamoto, S.,: Fuzzy Multisets and Their Generalizations. In: Calude, C.S., Paun, G., Rozenberg, G., Salomaa A. (eds.). Multiset Processing: Mathematical, Computer Science, and Molecular Computing Points of View. LNCS, vol. 2235, pp. 225--235. Springer, Heidelberg (2001)
24. Schmidt-Schauβ, M., Smolka, G.,: Attributive Concept Descriptions with Complements. Artificial Intelligence 48, 1--26 (1991)
25. Stoilos, G., Stamou, G., Pan, J.Z., Tzouvaras, V., Horrocks, I.,: Reasoning with Very Expressive Fuzzy Description Logics. Journal of Artificial Intelligence Research 30, 273--320 (2007)
26. Straccia, U.,: Reasoning within Fuzzy Description Logics. Journal of Artificial Intelligence Research 14, 137--166 (2001)
27. Syropoulos, A.,: Mathematics of Multisets. In: Calude, C.S., Paun, G., Rozenberg, G., Salomaa A. (eds.). Multiset Processing: Mathematical, Computer Science, and Molecular Computing Points of View. LNCS, vol. 2235, pp. 347--358. Springer, Heidelberg (2001)
28. Yager, R.,: On the Theory of Bags. International Journal of General Systems 13, 23--37 (1986)