RWTH
LTCS–Report

Aachen University of Technology
Research group for
Theoretical Computer Science

Computing the least common subsumer and the most specific concept in the presence of cyclic $\mathcal{ALN}$-concept descriptions

Franz Baader Ralf Küsters

LTCS-Report 98-06

An abridged version of this technical report has been published at KI'98.

RWTH Aachen
LuFg Theoretische Informatik
http://www-lti.informatik.rwth-aachen.de

Ahornstr. 55
52074 Aachen
Germany

Abstract

Computing least common subsumers (lcs) and most specific concepts (msc) are inference tasks that can be used to support the "bottom up" construction of knowledge bases for KR systems based on description logic. For the description logic $\mathcal{ALN}$, the msc need not always exist if one restricts the attention to acyclic concept descriptions. In this paper, we extend the notions lcs and msc to cyclic descriptions, and show how they can be computed. Our approach is based on the automata-theoretic characterizations of fixed-point semantics for cyclic terminologies developed in previous papers.

1 Introduction

Knowledge representation systems based on description logics (DL) can be used to describe the terminological knowledge of an application domain in a structured and formally well-understood way [3, 13]. Traditionally, the knowledge base of a DL system is built by first formalizing the relevant concepts of the domain (its terminology, stored in the so-called TBox) by *concept descriptions*, i.e., expressions that are built from atomic concepts (unary predicates) and atomic roles (binary predicates) using the concept constructors provided by the DL language. In a second step, the concept descriptions are used to specify properties of objects and individuals occurring in the domain (the world description, stored in the so-called ABox). DL systems provide their users with inference services that support both steps: classification of concepts and individuals and testing for consistency. Classification of concepts determines subconcept/superconcept relationships (called subsumption relationships) between the concepts of a given terminology, and thus allows one to structure the terminology in the form of a subsumption hierarchy. This hierarchy provides useful information on (implicit) connections between different concepts, and can thus be used to check (at least partially) whether the formal descriptions capture the intuitive meaning of the concepts. Classification of individuals (or objects) determines whether a given individual is always an instance of a certain concept (i.e., whether this instance relationship is implied by the descriptions of the individual and the concept). It thus provides useful information on the properties of an individual, and can again be used for checking the adequacy of the knowledge base with respect to the application domain it is supposed to describe. Finally, if a knowledge base is inconsistent (i.e., self-contradictory), then it is clear that a modeling error has occurred, and the knowledge base must be changed.

This traditional "top down" approach for constructing a DL knowledge base is not always adequate, though. On the one hand, it need not be clear from the outset which are the relevant concepts in a particular application. On the other hand, even if it is clear which (intuitive) concepts should be introduced, it is in general not easy to come up with formal definitions of these concepts within the available description language. For example, in one of our applications in chemical process engineering [4], the process engineers prefer to construct the knowledge base (which consists of descriptions of standard building blocks of process models, such as reactors) in the following "bottom up" fashion: first, they introduce several "typical" examples of the standard building block as individuals in the ABox, and then they generalize (the descriptions of) these individuals into a concept description that (a) has all the individuals as instances, and (b) is the most specific description satisfying property (a).

The present paper is concerned with developing inference services that can support this "bottom up" approach of building knowledge bases. We split the task of computing descriptions satisfying (a) and (b) from above into two subtasks: computing the most specific concept of a single ABox individual, and computing the least common subsumer of two concepts. The *most specific concept* (msc) of an individual b (the *least common subsumer* (lcs) of two concept descriptions A, B) is the most specific concept description C (expressible in the given description language) that has b as an instance (that subsumes both A and B). For sub-languages of the DL used by the system CLASSIC [5], both tasks have already been considered in the literature [6, 8, 7]. However, the algorithms described in these papers only compute approximations of the msc of an individual. In fact, for ABoxes with cyclic dependencies between individuals, the msc of a given individual need not exist, unless one allows for *cyclic concept descriptions* (i.e., concepts defined by cyclic TBoxes, interpreted with greatest fixed-point semantics). Once one allows for cyclic concept descriptions, the algorithm for computing the lcs must also be able to deal with these descriptions.

As a first solution to these problems, we consider cyclic concept descriptions in the language $\mathcal{ALN}$ (which allows for conjunctions, value restrictions, number restrictions, and atomic negations), and show how (1) the lcs of two such descriptions and (2) the msc of an ABox individual can be computed. In (2) we allow for cyclic descriptions in the ABox, and the msc may also be a cyclic description. Our approach is based on the known automata-theoretic characterizations of subsumption w.r.t. cyclic terminologies with greatest fixed-point semantics [2, 11].

2 Definitions and notations

In this section, we introduce the description language $\mathcal{ALN}$ as well as the notions msc and lcs more formally, and show how they can be generalized to cyclic concept descriptions.

Definition 1 ($\mathcal{ALN}$-concept descriptions).
$\mathcal{ALN}$-concept descriptions are formed from concept names and role names by means of the following syntax rules:

$$C, D \longrightarrow A \mid \neg A \mid C \sqcap D \mid \forall R.C \mid (\geq m\ R) \mid (\leq n\ R)$$

where A denotes a concept name, R a role name, C, D concept descriptions, m a positive integer, and n a non-negative integer. $\diamond$

The semantics of $\mathcal{ALN}$-concept descriptions is defined by introducing the notion of an interpretation. An *interpretation* I consists of a domain $dom(I)$ and a mapping assigning a subset A^I

of $dom(I)$ (the *extension of* A) to every concept name A, and a binary relation R^I over $dom(I)$ (the *extension of* R) to every role name R. This interpretation is extended to $\mathcal{ALN}$-concepts as defined in Table 1, where $R^I(d) := \{e \in dom(I) \mid (d, e) \in R^I\}$ denotes the set of R-*successors* of d in I.

syntax	semantics	name of construct
$\neg A$	$dom(I) \setminus A^I$	atomic negation
$C \sqcap D$	$C^I \cap D^I$	conjunction
$\forall R.C$	$\{d \in dom(I); R^I(d) \subseteq C^I\}$	value restriction
$(\geq m\ R)$	$\{d \in dom(I); \lvert R^I(d)\rvert \geq m\}$	at-least restriction
$(\leq n\ R)$	$\{d \in dom(I); \lvert R^I(d)\rvert \leq n\}$	at-most restriction

Table 1: Semantics of $\mathcal{ALN}$-concepts

In the following, we use $\bot$ to denote a concept description that is always interpreted by the empty set, such as $(\geq 2\ R) \sqcap (\leq 1\ R)$. In addition, we restrict our attention to the sub-language $\mathcal{FLN}$ of $\mathcal{ALN}$, which disallows atomic negation. In fact, as shown in [1], atomic negation can be simulated within $\mathcal{FLN}$, by using $(\leq 0\ R_A)$ in place of A and $(\geq 1\ R_A)$ in place of $\neg A$, where R_A is a new role name only used for this purpose.

Definition 2 (subsumption, lcs).
Let C, D, E be $\mathcal{FLN}$-concept descriptions.

1. C is *subsumed by* D ($C \sqsubseteq D$) iff $C^I \subseteq D^I$ holds for all interpretations I.
2. E is a *least common subsumer* (lcs) of C, D iff it satisfies
   - $C \sqsubseteq E$ and $D \sqsubseteq E$, and
   - E is the least $\mathcal{FLN}$-concept description with this property, i.e., if E' is an $\mathcal{FLN}$-concept description satisfying $C \sqsubseteq E'$ and $D \sqsubseteq E'$, then $E \sqsubseteq E'$.

$\diamond$

As shown in [6], the lcs of two $\mathcal{FLN}$-concept descriptions always exists, and it can be computed in polynomial time. Things become less rosy, however, if we consider the most specific concept of ABox individuals.

Definition 3 ($\mathcal{FLN}$-ABoxes).
An $\mathcal{FLN}$-*ABox* is a finite set of assertions of the form $R(a, b)$ (role assertion) or $C(a)$ (concept assertion), where a, b are individual names, R is a role name, and C is an $\mathcal{FLN}$-concept description. $\diamond$

In the presence of an ABox, an interpretation additionally assigns an element a^I of $dom(I)$ to each individual name a such that $a \neq b$ implies $a^I \neq b^I$ (unique name assumption). It is a *model* of the ABox $\mathcal{A}$ iff it satisfies $(a^I, b^I) \in R^I$ for all role assertions $R(a,b) \in \mathcal{A}$ and $a^I \in C^I$ for all concept assertions $C(a) \in \mathcal{A}$. An Abox is *consistent* if it has a model.

Definition 4 (instance, msc).
Let $\mathcal{A}$ be an $\mathcal{FLN}$-ABox, a an individual name in $\mathcal{A}$, and C an $\mathcal{FLN}$-concept description.

1. a is an *instance* of C w.r.t. $\mathcal{A}$ ($a \in_{\mathcal{A}} C$) iff $a^I \in C^I$ for all models I of $\mathcal{A}$.

2. C is the *most specific concept* for a in $\mathcal{A}$ iff $a \in_{\mathcal{A}} C$ and C is the least concept with this property, i.e., if C' is an $\mathcal{FLN}$-concept description satisfying $a \in_{\mathcal{A}} C'$, then $C \sqsubseteq C'$.

$\diamond$

The following example demonstrates that the msc need not exist if the ABox contains cyclic role assertions: in the ABox $\mathcal{A} := \{R(a,a), (\leq 1\ R)(a)\}$, the individual a does not have a most specific concept. In fact, it is easy to see that a is an instance of $\forall R. \cdots \forall R.((\leq 1\ R) \sqcap (\geq 1\ R))$ for chains of value restrictions of arbitrary length. Consequently, the msc cannot be expressed by a finite $\mathcal{FLN}$-concept description. However, the msc of a can be described by the recursively defined concept $A \doteq (\leq 1\ R) \sqcap (\geq 1\ R) \sqcap \forall R.A$, provided that this recursive definition is interpreted with greatest fixed-point semantics.

Definition 5 (cyclic $\mathcal{FLN}$-TBoxes).
An *$\mathcal{FLN}$-concept definition* is of the form $A \doteq C$, where A is a concept name and C an $\mathcal{FLN}$-concept description. An *$\mathcal{FLN}$-TBox* is a finite set of $\mathcal{FLN}$-concept definitions such that every concept name occurs at most once as left-hand side of a definition.[1]

The concept name A is a *defined concept* in the TBox $\mathcal{T}$ iff it occurs on the left-hand side of a definition in $\mathcal{T}$. Otherwise, A is called *primitive concept*. $\diamond$

The interpretation I is a model of the TBox $\mathcal{T}$ iff it satisfies $A^I = C^I$ for all concept definitions $A \doteq C \in \mathcal{T}$. It is well-known [12] that in the presence of cycles in the TBox, a given interpretation of the primitive concepts and roles (*primitive interpretation*) can have different extensions to a model of the TBox. An interpretation I is an extension of a primitive interpretation J if I coincide with J on the primitive concepts and roles. Additionally, I interprets the defined concepts of $\mathcal{T}$. The extensions of the defined concepts can be arranged in a tuple. These tuple can be ordered componentwise by set inclusion. The extensions of J to models of $\mathcal{T}$ are ordered by this ordering. The *gfp-semantics* chooses the greatest of the possible models as the gfp-model of the TBox.[2] Because this gfp-model is uniquely determined by the primitive

[1]Note that we do not prohibit cyclic dependencies between definitions.

[2]See [12, 2] for a more formal definition of the gfp-semantics, and for a discussion of other choices for the semantics of cyclic terminologies.

interpretation, the following definition of cyclic $\mathcal{FLN}$-concept descriptions and their semantics makes sense.

Definition 6 (cyclic $\mathcal{FLN}$-concept descriptions).
Assume that sets of *primitive* concept names N_P and of role names N_R are fixed. A *cyclic $\mathcal{FLN}$-concept description* $C = (A, \mathcal{T})$ is given by a defined concept A in a (possibly cyclic) $\mathcal{FLN}$-TBox $\mathcal{T}$ such that all the primitive concepts in $\mathcal{T}$ are elements of N_P and none of the defined concepts in $\mathcal{T}$ belongs to N_P. $\diamond$

In this context, an *interpretation* I assigns subsets of $dom(I)$ to elements of N_P and binary relations on $dom(I)$ to elements of N_R. For a given cyclic concept description $C = (A, \mathcal{T})$, the interpretation C^I of C in I is the set assigned to A by the unique extension of I to a gfp-model of T. This shows that, from a semantic point of view, cyclic concept descriptions C behave just like ordinary concept descriptions, i.e., a given interpretation I assigns a unique set $C^I \subseteq dom(I)$ to C. For this reason, the definition of subsumption and of the least common subsumer can be generalized to cyclic concept descriptions in the obvious way: just replace "$\mathcal{FLN}$-concept description" by "cyclic $\mathcal{FLN}$-concept description" in Definition 2. The same is true for the definitions of ABoxes, the instance relationship, and the most specific concept.

3 Computing the lcs of cyclic $\mathcal{FLN}$-concept descriptions

Both subsumption and the lcs of cyclic $\mathcal{FLN}$-concept descriptions can be computed using automata-theoretic characterizations of so-called value-restriction sets. For convenience, we abbreviate the concept description $\forall R_1.\forall R_2 \cdots \forall R_n.C$ $(n \geq 0)$ by $\forall R_1 \cdots R_n.C$, where $R_1 \dots R_n$ is a word over the alphabet N_R of all role names (i.e., $R_1 \dots R_n \in N_R^*$). For an interpretation I and a word $W = R_1 \dots R_n$, we define $W^I := R_1^I \circ \cdots \circ R_n^I$, where $\circ$ denotes the composition of binary relations.

Definition 7.
Let C be a cyclic $\mathcal{FLN}$-concept description and P a primitive concept name or a number restriction. Then the set $V_C(P) := \{W \in N_R^* \mid C \sqsubseteq \forall W.P\}$ is called the *value-restriction set* of C for P. $\diamond$

Even for acyclic descriptions, these value-restriction sets may be infinite. For example, for the (acyclic) description $\bot := (\geq 2\ R) \sqcap (\leq 1\ R)$ and an arbitrary primitive concept name P we have $V_\bot(P) = N_R^*$. The value-restriction sets can, however, be represented by regular languages over the alphabet N_R. To obtain these languages, the TBox $\mathcal{T}_C$ of a given cyclic $\mathcal{FLN}$-concept description C is translated into a semi-automaton $\mathcal{A}_{\mathcal{T}_C}$.

A *semi-automaton* is a triple (Σ, Q, Δ) where Σ is a finite alphabet, Q is a finite set of states, and $\Delta \subseteq Q \times \Sigma^* \times Q$ is a finite set of *word-transitions*.

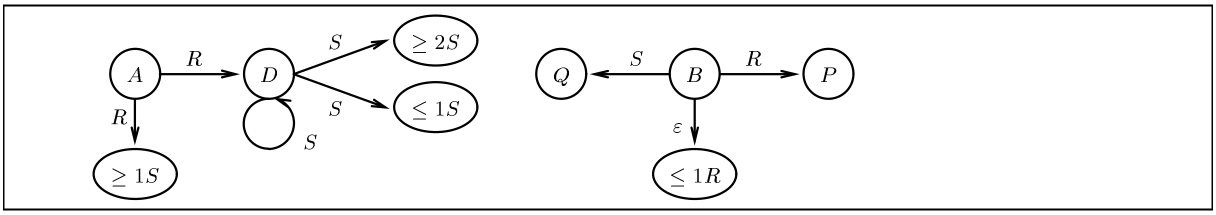

Figure 1: The automata corresponding to $\mathcal{T}_A$ and $\mathcal{T}_B$.

In order to define the semi-automaton $\mathcal{A}_{\mathcal{T}_C}$ of $\mathcal{T}_C$ we first have to normalize $\mathcal{T}_C$: $\mathcal{T}_C$ is called *normalized* if every right-hand side of a concept definition is a finite conjunction of concepts of the form $\forall W.A$ where W is a finite word over N_R^* and A is an atomic concept or a number-restriction.

Since $\forall R.(A \sqcap B)$ is equivalent to $\forall R.A \sqcap \forall R.B$, we can w.l.o.g. assume that a terminology is normalized.

Definition 8 (the semi-automaton of $\mathcal{T}_C$).
Let $\mathcal{T}_C$ be normalized. Then $\mathcal{A}_{\mathcal{T}_C} = (N_R, Q, \Delta)$ is the (non-deterministic) semi-automaton corresponding to $\mathcal{T}_C$ where the atomic concepts and number-restrictions of $\mathcal{T}_C$ are the states $q \in Q$ of $\mathcal{A}_{\mathcal{T}_C}$; a concept definition $A = \forall W_1.A_1 \sqcap \cdots \sqcap \forall W_k.A_k$ gives rise to k word-transitions where the transition from A to A_i is labeled by the word W_i. $\diamond$

Note that in a semi-automaton, word-transitions can easily be eliminated by replacing each of these transitions by a sequence of new introduced transitions (labeled with letters) using new states. Therefore, if needed, $\mathcal{A}_{\mathcal{T}_C}$ is (w.l.o.g.) supposed to be a *semi-automaton without word-transitions*, i.e., $\Delta \subseteq Q \times (N_R \cup \{\varepsilon\}) \times Q$ (see [2, 11] for details). For example, the TBoxes $\mathcal{T}_A$ and $\mathcal{T}_B$ defining the descriptions $C_A := (A, \mathcal{T}_A)$ and $C_B := (B, \mathcal{T}_B)$

$$\mathcal{T}_A: \quad A \doteq \forall R.D \sqcap \forall R.(\geq 1S) \qquad\qquad \mathcal{T}_B: \quad B \doteq (\leq 1R) \sqcap \forall R.P \sqcap \forall S.Q$$
$$D \doteq \forall S.D \sqcap \forall S.(\geq 2S) \sqcap \forall S.(\leq 1S)$$

give rise to the automata of Fig. 1. For a cyclic $\mathcal{FLN}$-concept description $C = (A, \mathcal{T})$ and a primitive concept or number restriction P, the language $L_C(P)$ is the set of all words labeling paths in the corresponding automaton from A to P. By definition, these languages are regular. In the example, we have, e.g., $L_{C_A}(\geq 2\ S) = RS^*S$ and $L_{C_B}(P) = \{R\}$.

It is easy to see that the inclusion $L_C(P) \subseteq V_C(P)$ always holds. However, since conflicting number restrictions can create inconsistencies (i.e., unsatisfiable sub-concepts), the inclusion in the other direction need not hold. Additionally, the set $V_C(P)$ may contain so-called C-excluding words:

Definition 9.
Let C be a cyclic $\mathcal{FLN}$-concept description. Then the set $E_C := \{W \in N_R^* \mid C \sqsubseteq \forall W.\bot\}$ is called the *set of C-excluding words.* $\diamond$

Obviously, if $W \in L_C(\leq m\ R) \cap L_C(\geq n\ R)$ for $m < n$, then W must belong to E_C. Also, since $(\leq 0\ R)$ is equivalent to $\forall R.\bot$, we know that $W \in L_C(\leq 0\ R)$ implies $WR \in E_C$. In addition, if W belongs to E_C, then $WU \in E_C$ for all words U. Finally, for $W \in E_C$, at-least restrictions can also force prefixes of W to belong to E_C. In our example (see Fig. 1), the word R belongs to E_{C_A} since $RS \in E_{C_A}$ and $R \in L_{C_A}(\geq 1\ S)$. Consequently, $E_{C_A} = R\{R, S\}^*$ and it is easy to see that $E_{C_B} = \emptyset$

A more formal characterization of E_C, which also shows that E_C is a regular language, can be found in [11]. To be more precise, a finite automaton that accepts E_C and is exponential in the size of the automaton corresponding to C can be constructed. The following characterization of value-restriction sets is an easy consequence of the results in [11]:

Theorem 10.
Let C be a cyclic $\mathcal{FLN}$-concept description. Then

1. $V_C(P) = L_C(P) \cup E_C$ for all primitive concepts P;
2. $V_C(\geq m\ R) = \bigcup_{\ell \geq m} L_C(\geq \ell\ R) \cup E_C$ for all at-least restrictions $(\geq m\ R)$;
3. $V_C(\leq n\ R) = \bigcup_{\ell \leq n} L_C(\leq \ell\ R) \cup E_C R^{-1}$ for all at-most restrictions $(\leq n\ R)$.[3]

$\square$

Consequently, these sets are regular, and finite automata accepting them can be constructed in time exponential in the size of the automaton corresponding to C.

Using the notion of value-restriction sets, the automata-theoretic characterization of subsumption of cyclic $\mathcal{FLN}$-concept descriptions provided in [11] can be formulated as follows:

Proposition 11.
Let C, D be cyclic $\mathcal{FLN}$-concept descriptions. Then $C \sqsubseteq D$ iff $L_D(P) \subseteq V_C(P)$ for all primitive concept names and number restrictions P. $\square$

As an easy consequence, we obtain the following characterization of the lcs of such descriptions:

Corollary 12.
Let C, D be cyclic $\mathcal{FLN}$-concept descriptions. Then the cyclic $\mathcal{FLN}$-concept description E is the lcs of C and D if $L_E(P) = V_C(P) \cap V_D(P)$ for all primitive concept names or number restrictions P. $\square$

[3]For a language L and a letter R, we define $LR^{-1} := \{W \mid WR \in L\}$.

Given automata for the (non-empty) value-restriction sets $V_C(P)$ and $V_D(P)$, it is easy to construct a cyclic $\mathcal{FLN}$-concept description E that satisfies this property (by simply translating the automata back into TBoxes). This shows that the lcs of two cyclic $\mathcal{FLN}$-concept descriptions can be computed in exponential time, and its size is at most exponential in the size of the input descriptions.

4 Inconsistency of an ABox

The aim of this section is to prove an automata theoretic characterization of inconsistency of an $\mathcal{FLN}$-ABox $\mathcal{A}$ with cyclic descriptions. From this characterization we derive decision algorithms and complexity results for inconsistency. First, we have to introduce some basic notions, which will be used even in the subsequent sections.

In the following, we let $\mathcal{A}$ be an arbitrary but fixed $\mathcal{FLN}$-ABox with cyclic concept descriptions, $I_{\mathcal{A}}$ the set of individuals in $\mathcal{A}$, $a, b \in I_{\mathcal{A}}$ individuals in $\mathcal{A}$ and P a primitive concept or a number-restriction. If I is a model of $\mathcal{A}$, $d, e \in dom(I)$, $V \in N_R^*$ where dV^Ie, then e is a *V-successor* of d and we refer to dV^Ie as path in I from d to e with label V.

In addition to the automata corresponding to the cyclic concept descriptions in $\mathcal{A}$, we need an *semi-automaton corresponding to* $\mathcal{A}$: the states of this automaton are the individual names occurring in $\mathcal{A}$, and the transitions are just the role assertions of $\mathcal{A}$, i.e., there is a transition with label R from a to b iff $R(a, b) \in \mathcal{A}$. For individual names a, b occurring in $\mathcal{A}$, the (regular) language $L_a(b)$ is the set of all words labeling paths from a to b in this automaton. We say that there is a *role chain in* $\mathcal{A}$ from the individual a to b with label U (aUb for short) iff $U \in L_a(b)$. Note, that $a\varepsilon a$ is also a role chain in $\mathcal{A}$. The set $R_{\mathcal{A}}(a) := \{b \mid R(a, b) \in \mathcal{A}\}$ denotes the set of R-successors of a in $\mathcal{A}$.

In the following we introduce the notions "predecessor restriction set" and "number conditions" which represent value-restrictions and number-restrictions that must be satisfied by individuals of $\mathcal{A}$.

Definition 13 (predecessor restriction sets).
The following sets are *predecessor restriction sets* for a (w.r.t. $\mathcal{A}$ and P):

$$L_a(P) \;:=\; \{W \in N_R^* \mid \text{there is a concept assertion } C(b) \in \mathcal{A}, \text{ a word } U \in L_b(a) \text{ such that } UW \in L_F(P)\}$$

$\diamond$

The meaning of these sets is stated in

Lemma 14.
Let $W \in N_R^*$. Then $W \in L_a(P)$ implies $a \in_{\mathcal{A}} \forall W.P$.

Proof. If $W \in L_a(P)$, then there is a concept assertion $F(f) \in \mathcal{A}$ and a word $U \in L_f(a)$ such that $UW \in L_F(P)$. Let $d \in dom(I)$ be an W-successor of a^I, i.e., a^IW^Id. Since I is an model of $\mathcal{A}$, it follows $f^I U a^I W^I d$ and $f^I \in F^I$. Now, $UW \in L_F(P)$ implies $d \in P^I$ (see Theorem 22, [10]). This shows $a^I \in (\forall W.P)^I$. $\square$

For number restrictions we are, in addition, interested in the maximum and minimum number restrictions which an individual and its successors has to satisfy.

Definition 15 (number conditions).
For $V \in N_R^*$ and $R \in N_R$ we define:

$$\begin{aligned} c_{a,V}^{\geq R} &:= max\{n \mid V \in L_a(\geq nR)\} \ (max(\emptyset) := 0), \\ c_{a,V}^{\leq R} &:= min\{n \mid V \in L_a(\leq nR)\} \ (min(\emptyset) := \infty). \end{aligned}$$

the *R-number conditions* for V-successors of a. $\diamond$

In every model of $\mathcal{A}$ the number of R-successors of V-successors of a has to satisfy these number conditions. Formally, this means:

Lemma 16.
Let I be a model of $\mathcal{A}$, $V \in N_R^*$, $d \in dom(I)$ where $a^I V^I d$. Then $c_{a,V}^{\geq R} \leq |R^I(d)| \leq c_{a,V}^{\leq R}$.

Proof. First we show $c_{a,V}^{\geq R} \leq |R^I(d)|$. We have to distinguish two cases:

- There is a non-negative integer n such that $V \in L_a(\geq nR)$ and $n = c_{a,V}^{\geq R}$. According to Lemma 14 it follows $a^I \in (\forall V.(\geq n\ R))^I$. But then, $a^I V^I d$ implies $d \in (\geq n\ R)^I$. This shows $c_{a,V}^{\geq R} \leq |R^I(d)|$.
- If $c_{a,V}^{\geq R} = 0$, then there is nothing to show.

The statement $c_{a,V}^{\leq R} \geq |R^I(d)|$ can be proved analogously. $\square$

Finally, we have to capture the fact that certain successors are required by an individual.

Definition 17 (require).
Let $V, V'(= R_1 \cdots R_n) \in N_R^*$. Then, we say VV' is *required by the individual a starting from V* if $c_{a,VR_1 \cdots R_{i-1}}^{\geq R_i} \geq 1$ for all $1 \leq i \leq n$. If $V = \varepsilon$, then we say V' is *directly required by the individual a*. We say that a word W is *required by the individual a* if there is V, V' and an individual b in $\mathcal{A}$ such that $W = VV'$, aVb is a role chain in $\mathcal{A}$, and V' is directly required by b. $\diamond$

The meaning of this definition is presented in the following lemma and can easily be proved using Lemma 16.

Lemma 18.
Let I be a model of $\mathcal{A}$, $d \in dom(I)$, $V, V' \in N_R^*$, and $a^I V d$. Then it holds: If VV' is required by a starting from V, then there is a V'-successor of d. If V is required by a, then there is a V-successor of a. $\square$

Now we are equipped to characterize inconsistency of an ABox. Intuitively, $\mathcal{A}$ is inconsistent if $\mathcal{A}$ requires at least one individual in every possible model for which conflicting number-restrictions are stated. In the sequel, we formally prove the following

Theorem 19 (Characterization of inconsistency).
The ABox $\mathcal{A}$ is inconsistent iff

1. there is an individual a in $\mathcal{A}$, $R \in N_R$ such that $|R_{\mathcal{A}}(a)| > c^{\leq R}_{a,\varepsilon}$, or
2. there is an individual a in $\mathcal{A}$, $V \in N_R^*$, and $R \in N_R$ such that V is directly required by a and $c^{\geq R}_{a,V} > c^{\leq R}_{a,V}$.

□

It is not hard to prove that this characterization is sufficient for the inconsistency of $\mathcal{A}$: Let I be a model of $\mathcal{A}$. Let a be an individual in $\mathcal{A}$ and $R \in N_R$. By Lemma 16 it follows $|R^I(a)| \leq c^{\leq R}_{a,\varepsilon}$. Thus, by the unique name assumption we have $|R_{\mathcal{A}}(a)| \leq c^{\leq R}_{a,\varepsilon}$. Furthermore, assuming that there is an individual a in $\mathcal{A}$ and a V directly required by a it follows by Lemma 18 that a has a V-successor d in I. By Lemma 16 it follows $c^{\geq R}_{a,V} \leq R^I(d) \leq c^{\leq R}_{a,V}$, and thus $c^{\geq R}_{a,V} \leq c^{\leq R}_{a,V}$.

If the right-hand side of the equivalence in Theorem 19 does not hold, then it is possible to construct a model of $\mathcal{A}$.

For this purpose we have to introduce some generalizations of predecessor restriction sets and number conditions. The predecessor restriction sets state (universal) value-restrictions which must hold for all models of $\mathcal{A}$. We will also need such sets for given role interpretations. A *role interpretation* of the ABox $\mathcal{A}$ consists of the domain $dom(J)$ and a mapping assigning a binary relation R^I to every role name $R \in N_R^*$ as well as an element a^I to every individual name in $\mathcal{A}$ such that $(a^I, b^I) \in R^I$ for every role assertion $R(a,b) \in \mathcal{A}$.

Definition 20 (predecessor restriction sets w.r.t. role interpretations).
Let J be a role interpretation and $d \in dom(J)$. The following set is called *predecessor restriction set* for d and P w.r.t. J (and $\mathcal{A}$):

$$L_{d,J}(P) \;:=\; \{W \in N_R^* \mid \text{ there is concept assertion } F(f) \in \mathcal{A}, \text{ a word } U \in N_R^* \text{ such that } UW \in L_F(P) \text{ and } f^J U^J d\},$$

◇

Analogously to Lemma 14 we can show

Lemma 21.
Let J be a role interpretation for $\mathcal{A}$, I a model of $\mathcal{A}$ such that I and J have the same domain and coincide on the interpretation of the roles. Finally, let $d \in dom(J)$. Then, for every $W \in N_R^*$ and every primitive interpretation or number-restriction P: $W \in L_{d,J}(P)$ implies $d^I \in (\forall W.P)^I$,
□

As above we are interested in number restrictions of individuals; now, not only individuals in $\mathcal{A}$, but more generally individuals of a role interpretation are considered.

Definition 22 (number conditions w.r.t. role interpretations).
Let J be a role interpretation, $d \in dom(J)$, and $R \in N_R$. Then, we define:

$$c^{\geq R}_{d,J} \;:=\; max\{n \mid \varepsilon \in L_{d,J}(\geq nR)\} \;(max(\emptyset) := 0),$$
$$c^{\leq R}_{d,J} \;:=\; min\{n \mid \varepsilon \in L_{d,J}(\leq nR)\} \;(min(\emptyset) := \infty).$$

the *R-number conditions* of d w.r.t. J. ◇

Using Lemma 21 the following lemma, which is an generalization of Lemma 16, can easily be verified.

Lemma 23.
Let J be a role interpretation for $\mathcal{A}$, I be a model of $\mathcal{A}$ such that I and J coincide on the interpretations of the roles, and let $d \in dom(J)$. Then: $c^{\geq R}_{d,J} \leq |R^I(d)| \leq c^{\leq R}_{d,J}$. □

Since every role interpretation J of $\mathcal{A}$ satisfies the role assertions in $\mathcal{A}$, we get the following relation of the notions "predecessor restriction sets" and "number conditions" on the one hand side and "predecessor restriction sets" and "number conditions" w.r.t. a role interpretation on the other hand side.

Lemma 24.
Let J be a role interpretation for $\mathcal{A}$, $V \in N_R^*$, and $n \geq 0$. Then:

- $L_a(\geq nR) \subseteq L_{a,J}(\geq nR)$;
- if $a^J V^J d$, then $c^{\geq R}_{a,V} \leq c^{\geq R}_{d,J}$;
- if $a^J V^J d$, then $c^{\leq R}_{a,V} \geq c^{\leq R}_{d,J}$.

□

With this definitions at hand we can specify the canonical model of $\mathcal{A}$.

Definition 25 (canonical model of $\mathcal{A}$).
We first define the *canonical role interpretation* J of $\mathcal{A}$ inductively.

J_0: $dom(J_0) := I_{\mathcal{A}}$; for all $R \in N_R$ let $R^{J_0} := \{(a,b) \mid R(a,b) \in \mathcal{A}\}$.

J_{i+1}: For all $R \in N_R$ and $d \in dom(J_i)$ where $(c^{\geq R}_{d,J_i} - |R^{J_i}(d)|) > 0$ the domain of J_i is extended by the new introduced, pairwise distinct individuals $d_1, \ldots, d_{c^{\geq R}_{d,J_i} - |R^J_i(d)|}$. Furthermore, these new individuals are added to the set of R-successors of d.

Now, the primitive interpretation J of $\mathcal{A}$ is defined by

$$\begin{aligned} dom(J) &:= \bigcup_{i\geq 0} dom(J_i) \\ R^J &:= \bigcup_{i\geq 0} R^{J_i} \\ P^J &:= \{d \in dom(J) \mid \varepsilon \in L_{d,J}(P)\}. \end{aligned}$$

The canonical model I of $\mathcal{A}$ is induced by J. ◇

The individuals in $dom(J) \setminus dom(J_0)$ are called *new elements of* J; the others are called *old*. An individual $e \in dom(J)$ is *generated in* J_k if $k = 0$ and $e \in dom(J_k)$ or if $e \in dom(J_k) \setminus dom(J_{k-1})$ for $k \geq 1$. Note that there is exactly one $k \geq 0$ such that e is generated in J_k. If $i < j$, then we say that the individuals generated in J_j are *generated later* than the individuals in J_i. Let $d, e \in dom(J)$ and $W \in N_R^*$. Then, we refer to $dW^J e$ as *new role chain* if apart from d this role chain contains only new elements. We summarize important properties of the canonical model in

Lemma 26.
Using the denotations in Definition 25 it holds:

1. Every new element d in J has only successors in J which are generated later than d.

2. For every $d \in dom(J)$ and $R \in N_R$ there is at most one $k \geq 1$ such that new generated elements in J_k are R-successors of d. More precise, if d is generated in J_i, $i \geq 0$, then $k = i + 1$.

3. Let a be an old individual, d be a new individual, and $W \in N_R^*$ such that $a^J W^J d$ is a new role chain. Then, if b is an old individual, $V \in N_R^*$ where $b^J V^J d$ is a new role chain, it follows $a = b$, $V = W$ and the role chains $a^J W^J d$ and $b^J V^J d$ are equal.

4. Let a be an old individual and $W \in N_R^*$. Then, we have: There is a $d \in dom(J)$ where $a^J W^J d$ iff W is required by a.

5. Let $d \in dom(J)$, $R \in N_R$. Then, it holds: If $|R^{J_0}(d)| \leq c_{d,J}^{\geq R}$, then $|R^J(d)| = c_{d,J}^{\geq R}$. The premise can only be invalid if d is an old element.

6. If the right-hand side in Theorem 19 is not satisfied, then I is a model of $\mathcal{A}$.

Proof. 1.: If d is generated in J_i, $i \geq 1$, then d has by definition no successors in J_i, thus, all successors of d are generated later.
2.: Let $d \in dom(J_i)$, $i \geq 0$, be generated in J_i and $R \in N_R$.

Claim: $c^{\geq R}_{d,J_i} = c^{\geq R}_{d,J_j}$ for all $j \geq i$

Proof of the claim: "$\leq$": Since J_j is an extension of J_i, it follows $L_{d,J_i}(\geq nR) \subseteq L_{d,J_j}(\geq nR)$. As an easy consequence we can conclude $c^{\geq R}_{d,J_i} \leq c^{\geq R}_{d,J_j}$.

"$\geq$": Let $n = c^{\geq R}_{d,J_j}$. This means, $\varepsilon \in L_{d,J_j}(\geq nR)$. According to the definition of $L_{d,J_j}(\geq nR)$ there is a concept assertion $F(f) \in \mathcal{A}$, and a word $U \in N_R^*$ such that $U \in L_F((\geq n\ R))$ and $f^{J_j} U^{J_j} d$. The role chain $f^{J_j} U^{J_j} d$ can only contain individuals in J_i; otherwise d is a successor of an individual which is generated later than d, which would be a contradiction to 1. Thus, $\varepsilon \in L_{d,J_i}(\geq nR)$, and hence, $c^{\geq R}_{d,J_i} \geq c^{\geq R}_{d,J_j}$. This completes the proof of the claim.

Now, we distinguish two cases:

- It holds $(c^{\geq R}_{d,J_i} - |R^{J_i}(d)|) > 0$. Then by construction, in J_{i+1} $(c^{\geq R}_{d,J_i} - |R^{J_i}(d)|)$ R-successors are generated for d such that $c^{\geq R}_{d,J_i} = |R^{J_{i+1}}(d)|$. By the claim it holds $c^{\geq R}_{d,J_i} = c^{\geq R}_{d,J_j}$ for all $j \geq i$. Consequently, in J_j, $j > i+1$, no new R-successors of d are generated.
- It holds $(c^{\geq R}_{d,J_i} - |R^{J_i}(d)|) \leq 0$. The claim implies $(c^{\geq R}_{d,J_j} - |R^{J_j}(d)|) \leq 0$ for all $j \geq i$. Thus, by construction no R-successors of d are generated.

3.: By construction of J, every new element has exactly one predecessor. This implies the claim in 3.

4.: "$\Leftarrow$" If W is required by a, then there are $V, V' \in N_R^*$ as well as an old element b such that aVb is a role chain in $\mathcal{A}$ and V' is directly required by b. Let $V' = R_1 \cdots R_n$. By induction over the length of V' the existence of an $d_i \in dom(J_i)$, $0 \leq i \leq n$, such that $b^{J_i}(R_1 \cdots R_i)^{J_i} d_i$ can easily be proved. As a consequence we have $a^J V^J b^J V'^J d$ where $d = d_n$.

"$\Rightarrow$" If d is an old element, then by 1 it follows that aWd is a role chain in $\mathcal{A}$, and thus, d is required by a.

Now, let d be a new element. Then, there are $V, V' \in N_R^*$, and an old element b such that $W = VV'$, aVb is a role chain in $\mathcal{A}$, and $b^J V'^J d$ is a new role chain. Let $V' = R_1 \cdots R_n$ as well as $d_1, \ldots, d_{n-1}$ new elements in J, $d_0 := b^J$, and $d_n := d$ such that $d_0 R_1^J d_1 R_2^J \cdots R_n^J d_n$ is the new role chain from d_0 to d_n. By the construction and 2. it holds $(c^{\geq R_{i+1}}_{d_i,J_i} - |R^{J_i}_{i+1}(d_i)|) > 0$ for all $0 \leq i < n$. This implies $c^{\geq R_{i+1}}_{d_i,J_i} \geq 1$ for all $0 \leq i < n$. Thus, for every $0 \leq i < n$ there is a concept assertion $F_i(f_i) \in \mathcal{A}$ and a word $U_i \in N_R^*$ such that $U_i \in L_{F_i}((\geq n_{i+1}\ R_{i+1}))$, $n_{i+1} \geq 1$, and $f_i^{J_i} U_i^{J_i} d_i$. By 3, there exists $V_i, V_i' \in N_R^*$ such that $U_i = V_i V_i'$ and $f_i^{J_i} V_i^{J_i} d_0 V_i'^{J_i} d_i$ where $d_0 V_i'^{J_i} d_i$ is the new role chain in J from d_0 to d_i with label $R_1 \cdots R_i$ and $f_i V_i D_0$ is a role chain in $\mathcal{A}$. This implies $c^{\geq R_{i+1}}_{b,R_1\cdots R_i} \geq 1$ for all $0 \leq i < n$. Thus, V' is directly required by b. This shows, that $W = VV'$ is required by a.

5.: Let d be an element generated in J_i, $i \geq 0$.

- Let d be an old element in J such that $|R^{J_0}(d)| \leq c^{\geq R}_{d,J}$. According to the claim in the proof of 2, it holds $c^{\geq R}_{d,J_0} = c^{\geq R}_{d,J}$. We know $(c^{\geq R}_{d,J_0} - |R^{J_0}(d)|) \geq 0$. In the case of

$(c^{\geq R}_{d,J_0} - |R^{J_0}(d)|) = 0$ by 2 no R-successors for d are generated. Thus, $|R^J(d)| = c^{\geq R}_{d,J}$. In the case of $(c^{\geq R}_{d,J_0} - |R^{J_0}(d)|) > 0$ R-successors of d are generated in J_1 such that $c^{\geq R}_{d,J_0} = |R^{J_1}(d)|$. By 2 and the claim in 2 it follows $|R^{J_1}(d)| = |R^J(d)| = c^{\geq R}_{d,J}$.

- Let d be a new element in J. By construction d has no successors in J_i, $i =\geq 1$, thus $0 = |R^{J_0}(d)| \leq c^{\geq R}_{d,J}$. By 2 only in J_{i+1} R-successor are generated for d. By the claim in the proof of 2, we have $c^{\geq R}_{d,J_i} = c^{\geq R}_{d,J}$, and thus, $0 = |R^{J_i}(d)| \leq c^{\geq R}_{d,J_i}$. Now, analogously to 1 we can show $|R^J(d)| = c^{\geq R}_{d,J}$.

6.: Assume that the right-hand side of Theorem 19 does not hold. Let $C(a) \in \mathcal{A}$. It is to show: $a^I \in C^I$. We use Theorem 22, [10] and distinguish the following cases:

(P1) Let $W \in N_R^*$, $d \in dom(I)$, P a primitive concept such that $a^I W^I d$ and $W \in L_C(P)$. Thus, $\varepsilon \in L_{d,J}(P)$. This means $d \in P^I$.

(P2) Let $W \in N_R^*$, $d \in dom(I)$, $(\geq n\ R)$ an at-least restriction such that $a^I W^I d$ and $W \in L_C((\geq n\ R))$. By 5 it follows $|R^J(d)| \geq c^{\geq R}_{d,J}$. Since $C(a) \in \mathcal{A}$, $a^I W^I d$, and $W \in L_C((\geq n\ R))$, we know $\varepsilon \in L_{d,J}(\geq nR)$. Hence, $c^{\geq R}_{d,J} \geq n$, and thus $|R^I(d)| \geq n$.

(P3) Let $W \in N_R^*$, $d \in dom(I)$, $(\leq n\ R)$ a at-most restriction such that $a^I W^I d$ and $W \in L_C((\leq n\ R))$. Assume: $d \notin (\leq n\ R)^I$. Then, $|R^I(d)| > n$. We distinguish two cases:

- It holds $d \in dom(J_0)$ and $|R^{J_0}(d)| > n$. By 1. we know that aWd is a role chain in $\mathcal{A}$. Consequently, it follows $c^{\leq R}_{d,\varepsilon} \leq n$. This is a contradiction to the assumption that 1. in Theorem 19 does not hold.
- It holds $d \in dom(J)$ and $|R^{J_0}(d)| \leq n$. Since $|R^J(d)| > n$, we know $m := c^{\geq R}_{d,J} = |R^J(d)| > n$. Further by 4., W is required by a. Consequently, there is an old individual b as well as $V, V' \in N_R^*$ such that $W = VV'$, aVb is a role chain in $\mathcal{A}$, $b^I V'^I d$ is a new role chain in J, and V' is directly required by b. Because of $m > 0$ there is a concept assertion $F(f) \in \mathcal{A}$ and a word $U \in L_F((\geq m\ R))$ such that $f^J U^J d$. If d is an old individual it follows $b = d$, $V' = \varepsilon$, and fUd is a role chain in $\mathcal{A}$ (3). Thus, $c^{\geq R}_{b,V'} \geq m > n$. If d is a new individual, then by 3 there is an U' such that $U = U'V'$ and $f^J U'^J b^J V'^J d$. Thus, we have $c^{\geq R}_{b,V'} \geq m > n$ as well. On the other hand aVb role chain in $\mathcal{A}$, $b^J V'^J d$, and $W \in L_C(\leq n\ R)$ imply $c^{\leq R}_{b,V'} \leq n < m$. This yields $c^{\leq R}_{b,V'} < c^{\geq R}_{b,V'}$ where V' is directly required by b. This is a contradiction to the assumption that 2 in Theorem 19 does not hold.

Thus, we have shown $d \in (\leq n\ R)^I$.

□

The only-if direction of Theorem 19 is an immediate consequence of Lemma 26, 6.

Finally, we prove the complexity upper and lower bound of inconsistency using Theorem 19.

In [10] it has been shown that inconsistency of concepts w.r.t. cyclic $\mathcal{FLN}$-terminologies is PSPACE-complete. Let T be a cyclic $\mathcal{FLN}$-terminology and C be an atomic concept in T. If the ABox $\mathcal{A}$ only consists of the concept assertion $C(a)$ where C is a cyclic $\mathcal{FLN}$-concept description defined by T, then C is consistent w.r.t. T iff $\mathcal{A}$ is consistent. This reduction shows that inconsistency of Aboxes with cyclic $\mathcal{FLN}$-concept descriptions is PSPACE-hard. In order to proof the complexity upper bound, we introduce the notion p-exclusion set. Therefore, additional definitions are necessary.

Every (cyclic) $\mathcal{FLN}$-concept description $C(a) \in \mathcal{A}$ can be defined by a (cyclic) $\mathcal{FLN}$-TBox.[4] W.l.o.g. we assume that the concept descriptions occurring on the right-hand side of concept definitions are conjunctions of concepts of the form A or $\forall R..A$ where A denotes a atomic concepts or a number-restriction, i.e., we do not allow for nested value-restrictions. We refer to the set of all atomic concepts and number-restrictions occurring in these TBoxes as $\mathcal{C}$. The next notion represents a set of all atomic concepts and number-restrictions that must be satisfied by W-successors of instances of concepts in $F \subseteq \mathcal{C}$:

$$next_\varepsilon(F, W) := \{A \in \mathcal{C} \mid \text{ there is a } B \in F \text{ such that } W \in L_B(A)\}.$$

Similar to the definition in [10], p-exclusion sets contain atomic concepts and number-restrictions that require successors leading to conflicting number-restrictions.

Definition 27 (p-exclusion set).
The set $F_0 \subseteq \mathcal{C}$ is called *p-exclusion set* if the following holds: There is a word $R_1 \cdots R_n \in N_R^*$, conflicting number-restrictions $(\geq l\ R)$ and $(\leq r\ R)$, $l > r$, and for all $1 \leq i \leq n$ there are integers $m_i \geq 1$ such that for $F_i := next_\varepsilon(F_0, R_1 \cdots R_i)$, $1 \leq i \leq n$, it holds that $(\geq m_i\ R_i) \in F_{i-1}$ for all $1 \leq i \leq n$ and $(\geq l\ R)$, $(\leq r\ R) \in F_n$. We denote the set of all p-exclusion sets $\mathcal{E}$. ◇

As shown in [10], for a given set $F \subseteq \mathcal{C}$ it is decidable if F is an p-exclusion set using polynomial space. So as to apply this result for deciding inconsistency of ABoxes we have to collect the set of atomic concepts and number-restriction in $\mathcal{C}$ that has to be satisfied by an individual.

Definition 28 (initial state of an individual).
The set

$$q_a := \bigcup_{\substack{F(f) \in \mathcal{A} \\ A \in \mathcal{C} \\ L_f(a) \cap L_F(A) \neq \emptyset}} \{A\}$$

is called the *initial set* of the individual a in $\mathcal{A}$. ◇

[4] In fact, all such concept descriptions can be presented in one TBox.

It can easily be verified that q_a is computable using polynomial space. In the following "algorithmic" characterization of inconsistency the initial set q_a of a takes the place of ε-$closure(A)$ in the characterization of inconsistency for the atomic concept A in a terminology T (see [10]).

Proposition 29.
The ABox $\mathcal{A}$ with cyclic $\mathcal{FLN}$-concept descriptions is inconsistent iff

1. there is an individual a in $\mathcal{A}$, $R \in \Sigma$, and $(\leq n\ R) \in q_a$ such that $|R_{\mathcal{A}}(a)| > n$, or
2. there is an individual a such that $q_a \in \mathcal{E}$.

Proof. "$\Rightarrow$" If $\mathcal{A}$ is inconsistent, then by Theorem 19 we have to distinguish two cases:

1. There is an individual a in $\mathcal{A}$, $R \in \Sigma$ such that $|R_{\mathcal{A}}(a)| > c_{a,\varepsilon}^{\leq R}$. Thus, there is an concept assertion $F(f) \in \mathcal{A}$, a word $X \in L_f(a)$, and an at-most restriction $(\leq n\ R)$ such that $X \in L_F((\leq n\ R))$ and $|R_{\mathcal{A}}(a)| > n$. Consequently, $(\leq n\ R) \in q_a$. This, implies 1. of the claim.
2. There is an individual a in $\mathcal{A}$, $V \in N_R^*$, and $R \in N_R$ such that V is directly required by a and $c_{a,V}^{\geq R} > c_{a,V}^{\leq R}$. Let $V = R_1 \cdots R_n$. Thus, there are a concept assertion $F_i(f_i) \in \mathcal{A}$, words $W_i \in L_{f_i}(a)$, and numbers $n_i \geq 1$ such that $W_i R_1 \cdots R_{i-1} \in L_{F_i}((\geq n_i\ R_i))$ for all $1 \leq i \leq n$. Consequently, since we do not allow for nested value-restrictions in the TBoxes corresponding to cyclic concept descriptions, there are $A_i \in \mathcal{C}$ such that $W_i \in L_{F_i}(A_i)$ and $R_1 \cdots R_{i-1} \in L_{A_i}((\geq n_i\ R_i))$ for all $1 \leq i \leq n$. Hence, $A_1, \ldots, A_n \in q_a$ and $(\geq n_i\ R_i) \in next_\varepsilon(q_a, R_1 \cdots R_{i-1})$. Additionally, there is a concept assertion $F(f) \in \mathcal{A}$, a word $W \in L_f(a)$, and a number $l = c_{a,V}^{\geq R}$ such that $WV \in L_F((\geq l\ R))$. As above there is a atomic concept or a number-restriction $A \in \mathcal{C}$ such that $W \in L_F(A)$ and $V \in L_A((\geq l\ R))$. Hence, $A \in q_a$ and $(\geq l\ R) \in next_\varepsilon(q_a, V)$. Analogously for $r = c_{a,V}^{\leq R}(< l)$ it can be shown $(\leq r\ R) \in next_\varepsilon(q_a, V)$. This shows 2. of the claim, i.e., q_a is an p-exclusion set.

"$\Leftarrow$" We distinguish two cases.

3. There is an individual a in $\mathcal{A}$, $R \in \Sigma$, and $(\leq n\ R) \in q_a$ such that $|R_{\mathcal{A}}(a)| > n$. By definition of q_a there is a concept assertion $F(f) \in \mathcal{A}$, a word $W \in L_f(a)$ such that $W \in L_F((\leq n\ R))$. This implies $c_{a,\varepsilon}^{\leq R} \leq n$, thus, $|R_{\mathcal{A}}(a)| > c_{a,\varepsilon}^{\leq R}$. By Theorem 19, 1. this means that $\mathcal{A}$ is inconsistent.
4. There is an individual a such that $q_a \in \mathcal{E}$. By definition of $\mathcal{E}$ there is a word $V = R_1 \cdots R_n \in N_R^*$, conflicting number-restrictions $(\geq l\ R)$ and $(\leq r\ R)$, $l > r$, and for all $1 \leq i \leq n$ there are integers $m_i \geq 1$ such that for $F_0 := q_a$ and $F_i := next_\varepsilon(F_0, R_1 \cdots R_i)$, $1 \leq i \leq n$, it holds that $(\geq m_i\ R_i) \in F_{i-1}$ for all $1 \leq i \leq n$ and $(\geq l\ R), (\leq r\ R) \in F_n$. Consequently, there are $A_1, \ldots, A_n$ as well as A^l, A^r in F_0 such that $R_1 \cdots R_{i-1} \in L_{A_i}((\geq m_i\ R_i))$ for all $1 \leq i \leq n$ as well as $V \in L_{A^l}((\geq l\ R))$ and $V \in L_{A^r}((\leq r\ R))$. By definition of q_a there are concept

assertions $F_i(f_i)$, and words $W_i \in L_{f_i}(a)$ for all $1 \leq i \leq n$ as well as concept assertions $F^l(f^l)$, $F^r(f^r)$, and words $W^l \in L_{f^l}(a)$, $W^r \in L_{f^r}(a)$ such that, and $W_i \in L_{F_i}(A_i)$ as well as $W^l \in L_{F^l}(A^l)$ and $W^r \in L_{F^r}(q^r)$. This means, $W_i R_1 \cdots R_{i-1} \in L_{F_i}((\geq m_i\ R_i))$ for all $1 \leq i \leq n$ as well as $W^l V \in L_{F^l}((\geq l\ R))$ and $W^r V \in L_{F^r}((\leq r\ R))$. Consequently, V is directly required by a and $c^{\geq R}_{a,V} \geq l > r \geq c^{\leq R}_{a,V}$. By Theorem 19, 2. it follows that $\mathcal{A}$ is inconsistent.

□

Since q_a is computable using polynomial space and since $q_a \overset{?}{\in} \mathcal{E}$ can be decided using polynomial space, by Proposition 29 it is not hard to see that inconsistency can be decided using polynomial space. To sum up, we have shown the following complexity result:

Corollary 30 (complexity of inconsistency).
Inconsistency of ABoxes with cyclic $\mathcal{FLN}$-concept descriptions is PSPACE-complete. □

5 Instance and most specific concept

In this section, we characterize instance and the msc of an individual specified in a cyclic ABox $\mathcal{A}$ with cyclic $\mathcal{FLN}$-concept descriptions. The key of these characterizations is to describe the set of universal value-restrictions which are satisfied by an individual. For that purpose, we first describe such sets.

5.1 Value-restriction sets of individuals

In this section, we answer the question which value-restrictions of the form $\forall W.P$ are satisfied by an individual name a occurring $\mathcal{A}$ where P is a primitive concept or a number-restriction. As main result it is shown that value-restriction sets are regular. Formally, these sets are defined as follows:

Definition 31 (value-restriction sets).
For a primitive concept P or a number-restriction P *value-restrictions sets* are defined as follows:

$$V_a(P) \quad := \quad \{W \in N_R^* \mid a \in_{\mathcal{A}} \forall W.P\}.$$

◇

By Lemma 14 it follows $L_a(P) \subseteq V_a(P)$. The language $L_a(P)$ corresponds to $L_C(P)$ in Section 3. Furthermore, as in Section 3 the inclusion relationship of $L_a(P)$ and $V_a(P)$ may be strict since there may exist a-excluding words W, i.e., $a \in_{\mathcal{A}} \forall W.\bot$, and thus, $a \in_{\mathcal{A}} \forall W.P$. We say that an a-excluding word that is induced by predecessors of a, i.e., by predecessor restriction sets of a, p-excludes a. Formally, p-exclusion is defined as follows:

Definition 32 (p-exclusion).
The word $W \in N_R^*$ *p-excludes* the individual a if

1. there exist $V, V' \in N_R^*$, $R \in N_R$ such that V is a prefix of W, VV' is required by a starting from V, and $c_{a,VV'}^{\geq R} > c_{a,VV'}^{\leq R}$; or

2. there is a prefix VR of W, $V \in N_R^*$, $R \in N_R$, such that $c_{a,V}^{\leq R} = 0$.

We refer to the set of p-exclusion words of a as E_a. $\diamond$

The next lemma shows that these words are in fact a-excluding words.

Lemma 33.
If $W \in E_a$, then $a \in_{\mathcal{A}} (\forall W.\bot)$.

Proof. Let $W \in E_a$ and $V, V' \in N_R^*$ specified as in Definition 32, 1. Assume that I is a model of $\mathcal{A}$, $d \in dom(I)$ such that $a^I W^I d$. Consequently, there is an individual e such that $a^I V^I e$. By Lemma 18 we know that there exists an V'-successor f of e. According to Lemma 16 it follows $c_{a,VV'}^{\geq R} \leq R^I(e) \leq c_{a,VV'}^{\leq R}$. This is a contradiction to $c_{a,VV'}^{\geq R} > c_{a,VV'}^{\leq R}$.

Now let $W \in E_a$, VR prefix of W where $V \in N_R^*$, $R \in N_R$, and $c_{a,V}^{\leq R} = 0$. Assume that I is a model of $\mathcal{A}$, $d \in dom(I)$ such that $A^I W^I d$. Consequently, there are individuals e, f such that $A^I V^I e R^I f$. Now Lemma 16 yields the contradiction $1 \leq R^I(f) \leq c_{a,V}^{\leq R} = 0$. $\square$

Using that $W \in (E_a \cdot R^{-1})$ implies $a \in_{\mathcal{A}} \forall W.(\leq 0\ R)$ we have as an easy consequence of Lemma 33 and Lemma 14 the following

Lemma 34.
For all primitive concepts P, at-least restrictions $(\geq n\ R)$, at-most restrictions $(\leq n\ R)$, individuals a in $\mathcal{A}$, and words $W \in N_R^*$ it holds that:

- if $W \in L_a(P) \cup E_a$, then $a \in_{\mathcal{A}} \forall W.P$;
- if $W \in L_a(\geq nR) \cup E_a$, then $a \in_{\mathcal{A}} \forall W.(\geq n\ R)$; and
- if $W \in L_a(\leq nR) \cup (E_a \cdot R^{-1})$, then $a \in_{\mathcal{A}} \forall W.(\leq n\ R)$.

$\square$

In general, the only-if directions in Lemma 34 are not true. Intuitively, this can be explained as follows: The definition of predecessor restriction sets $L_a(P)$ and thus of the p-exclusion set only takes into account value restrictions that come from predecessors of a. At-most restrictions in the ABox can, however, also require the propagation of value restrictions from successors of a back to a.

Let us first illustrate this phenomenon by a simple example. Assume that the ABox $\mathcal{A}$ consists of the following assertions:

$$R(a,b), \quad (\leq 1\ R)(a), \quad (\forall S.P)(b).$$

It is easy to see that $RS \notin L_a(P) \cup E_a$. However, $(\leq 1\ R)(a)$ ensures that, in any model I of $\mathcal{A}$, b^I is the only R^I-successor of a^I. Consequently, all $(RS)^I$-successors of a^I are S^I-successors of b^I, and thus $b^I \in (\forall S.P)^I$ implies $a^I \in (\forall RS.P)^I$. This shows that $RS \in V_a(P)$.

More generally, this problem occurs if concept assertions involving at-most restrictions in the ABox force role chains to use role assertions explicitly present in the ABox. In the example, we were forced to use the assertion $R(a,b)$ when going from a^I to an $(RS)^I$-successor of a^I. As a slightly more complex example, we assume that the ABox $\mathcal{A}$ contains the assertions

$$R(a,b), \quad R(a,c), \quad S(b,d), \quad (\leq 2\ R)(a), \quad (\forall R.(\leq 1\ S))(a),$$

and that $S \in L_c(P)$ and $\varepsilon \in L_d(P)$, where ε denotes the empty word. In a model I of $\mathcal{A}$, any $(RS)^I$-successor x of a^I is either equal to d^I or an S^I-successor of c^I. In the first case, $\varepsilon \in L_d(P)$ implies $x \in P^I$, and in the second case $S \in L_c(P)$ does the same. Consequently, we have $RS \in V_a(P)$, even though $RS \notin L_a(P) \cup E_a$. Here, we are forced to use either the assertions $R(a,b)$ and $R(b,d)$ or the assertion $R(a,c)$ when going from a^I to one of its $(RS)^I$-successors. Since in both cases the obtained successor must belong to P^I, a restriction on P must be propagated back to a from the successors of a.

Unfortunately, it is not yet clear how to give a *direct* characterization (as a regular language) of $V_a(P)$ that is based on an appropriate characterization of the set of words in $V_a(P) \setminus (L_a(P) \cup E_a)$ that come from this "backward propagation." Instead, we will describe the complement of $V_a(P)$ as a regular language. Since the class of regular languages is closed under complement, this also shows that $V_a(P)$ is regular.

In the example we have had words (such as RS) with prefixes (R in the examples) that must follow role chains of $\mathcal{A}$ in every model of $\mathcal{A}$. The words defined in the next sets do not have such prefixes.

$$N_a \quad := \quad \{\varepsilon\} \cup \bigcup_{\substack{S \in N_R \\ |S_{\mathcal{A}}(a)| < c^{\leq S}_{a,\varepsilon}}} S \cdot N_R^*,$$

$$N_a(\geq n\ R) \quad := \quad \begin{cases} N_a \setminus \{\varepsilon\} & ; \quad |R_{\mathcal{A}}(a)| \geq n \\ N_a & ; \quad \text{otherwise.} \end{cases}$$

Intuitively, a word of the form SU belongs to N_a if at-most restrictions in the ABox do not force all S-successors of a to be reached using role assertions explicitly present in the ABox. For at-least restrictions $(\geq n\ R)$ it depends on the restriction if the empty word belongs to $N_a(\geq n\ R)$.

Using these sets, value-restriction sets can be described as follows:

Theorem 35 (value-restriction sets).
Let b be an individual name in the *consistent* ABox $\mathcal{A}$, P a primitive concept, $(\geq n\ R)$ an at-least restriction, and $(\leq n\ R)$ an at-most restriction. Then

$$
\begin{aligned}
V_b(P) &= \overline{\bigcup_{c\in I_{\mathcal{A}}} L_b(c)\cdot(N_c\cap\overline{L_c(P)\cup E_c})}\\
V_b(\geq n\ R) &= \overline{\bigcup_{c\in I_{\mathcal{A}}} L_b(c)\cdot(N_c(\geq n\ R)\cap\overline{\bigcup_{m\geq n} L_c(\geq mR)\cup E_c})}\\
V_b(\leq n\ R) &= \overline{\bigcup_{c\in I_{\mathcal{A}}} L_b(c)\cdot(N_c\cap\overline{\bigcup_{m\leq n} L_c(\leq mR)\cup(E_c\cdot R^{-1})})}
\end{aligned}
$$

□

The hardest part of the proof of Theorem 35 is to show that value-restriction sets are contained in the sets of the right-hand side of the equations. But first, we proof the reverse inclusion relationship. We distinguish three cases:

1. Let $W\in V_b(P)$, I a model of $\mathcal{A}$, $d\in dom(I)$ such that b^IW^Id. We have to show $d^I\in P^I$. By the definition of $V_b(P)$ it follows for every $c\in I_{\mathcal{A}}$, $V,V'\in N_R^*$ where $W=VV'$, bVc role chain in $\mathcal{A}$, and $V'\in N_c$ that $V'\in L_c(P)\cup E_c$. Because of b^IW^Id there exist $c\in I_{\mathcal{A}}$, $V,V'\in N_R^*$ such that $b^IV^Ic^IV'^Id$, and $W=VV'$, bVc is a role chain in $\mathcal{A}$. Furthermore, if $V'\in S\cdot N_R^*$ we suppose that $|S^I(c)|>|S_{\mathcal{A}}(c)|$. By Lemma 16 it follows $c^{\geq S}_{\overline{c},\varepsilon}\leq|S^I(c^I)|\leq c^{\leq S}_{\overline{c},\varepsilon}$. Thus, $|S_{\mathcal{A}}(c)|<c^{\leq S}_{\overline{c},\varepsilon}$. This implies $V'\in N_c$. If such an V' does not exist, then we can assume V' to be the empty word. This yields $V'\in N_c$ as well. Now, by the assumption it follows $V'\in L_c(P)\cup E_c$, and Lemma 34 shows $d\in P^I$.

2. Let $W\in V_b(\geq l\ R)$, I a model of $\mathcal{A}$, $d\in dom(I)$ such that b^IW^Id. We have to show $d^I\in(\geq l\ R)^I$. By definition of $V_b(\geq l\ R)$ it follows for every $c\in I_{\mathcal{A}}$, $V,V'\in N_R^*$ where $W=VV'$, bVc role chain in $\mathcal{A}$, and $V'\in N_c(\geq l\ R)$ that $V'\in\bigcup_{r\geq l}L_c(\geq rR)\cup E_c$. Because of b^IW^Id there exist $c\in I_{\mathcal{A}}$, $V,V'\in N_R^*$ such that $b^IV^Ic^IV'^Id$, and $W=VV'$, bVc is a role chain in $\mathcal{A}$. Furthermore, if $V'\in S\cdot N_R^*$ for an $S\in N_R$ we suppose that $|S^I(c)|>|S_{\mathcal{A}}(c)|$. By Lemma 16 it follows $c^{\geq S}_{\overline{c},\varepsilon}\leq|S^I(c^I)|\leq c^{\leq S}_{\overline{c},\varepsilon}$. Thus, $|S_{\mathcal{A}}(c)|<c^{\leq S}_{\overline{c},\varepsilon}$. This implies $V'\in N_c(\geq l\ R)$, and thus, $V'\in\bigcup_{r\geq l}L_c(\geq rR)\cup E_c$. If such an V' does not exist, then we can assume V' to be the empty word. If $V'\in N_c(\geq l\ R)$, then again $V'\in\bigcup_{r\geq l}L_c(\geq rR)\cup E_c$ and Lemma 34 shows $d\in(\geq l\ R)^I$. If $V'\notin N_c(\geq l\ R)$, then by definition it follows $|R_{\mathcal{A}}(d)|\geq l$. Again, because of the unique name assumption this implies $d\in(\geq l\ R)^I$.

3. Let $W\in V_b(\leq l\ R)$, I a model of $\mathcal{A}$, $d\in dom(I)$ such that b^IW^Id. Analogously to 1. one can show $V'\in\bigcup_{r\leq l}L_c(\leq rR)\cup(E_c\cdot R^{-1})$. Thus, by Lemma 34 we have $d\in(\leq l\ R)^I$.

This shows that the expressions on the right-hand side of the equations of Theorem 35 only contain elements of the value-restriction sets.

In order to show the reverse inclusion relationship, we need a model of $\mathcal{A}$ such that if a word is not an element of a value-restriction set, the individual b is not contained in the extension of the corresponding value-restriction. The idea is to extend the canonical model I of $\mathcal{A}$ to $I(c, V')$ (for c and V' see the above proof of the if-direction of Theorem 35) such that b has a new role chain leading to c in $\mathcal{A}$ (labeled with V) and then leading from c labeled with V' to a new element which is, e.g., not contained in the extension of P. In case of at-most restrictions $(\leq l\ R)$ we further extend the path from c labeled with V' in $I(c, V')$ by $l+1$ R-successors to the model $I(c, V', R, l+1)$. Note that the path from c to the new element d labeled with V' is a new role chain, i.e., it contains no individuals of $\mathcal{A}$ beside c. This is crucial since in this context the value-restrictions that are satisfied by d are determined only by the predecessor restriction sets of c and E_c, i.e., "backward propagation" is avoided.

Definition 36 (extended canonical model).
Let a be an individual in $\mathcal{A}$, $W \in N_R^*$, $R \in N_R$, and r a non-negative integer. Let J denote the canonical primitive interpretation of $\mathcal{A}$ and I the canonical model of $\mathcal{A}$. We first define the *extended canonical primitive interpretation* $J' = J(a, W)$ and $J' = J(a, W, R, r)$, $r > 0$, of $\mathcal{A}$. This interpretation is defined inductively as follows where the interpretation of the primitive concepts are specified later on.

J_0: Let $U \in N_R^*$ be a maximal prefix of W such that there is a $d_1 \in dom(I)$ where $a^I U^I d_1$ is a new role chain in I. Let $V = R_1 \cdots R_n \in N_R^*$, $n \geq 0$, with $W = UV$. Furthermore, let $d_2, \ldots, d_{n+1} \notin dom(I)$ be new individuals. If $r - |R^I(d_{n+1})| \geq 0$, then let $k := r - |R^I(d_{n+1})|$, otherwise $k := 0$. Finally, also let $f_1, \ldots, f_k \notin dom(I) \cup \{d_2, \ldots, d_{n+1}\}$ be new individuals. Now, J_0 is obtained by extending the domain of J by $d_2, \ldots, d_{n+1}, f_1, \ldots, f_k$ as well as adding (d_i, d_{i+1}) to the extension of R_i for all $1 \leq i \leq n$ and adding $(d_{n+1}, f_1), \ldots, (d_{n+1}, f_k)$ to the extension of R.

J_{i+1}: For all $R \in N_R$ and $d \in dom(J_i)$ where $(c^{\geq R}_{d,J_i} - |R^{J_i}(d)|) > 0$ the domain of J_i is extended by the new introduced, pairwise distinct individuals $e_1, \ldots, e_{c^{\geq R}_{d,J_i} - |R^{J_i}(d)|}$. Furthermore, these new individuals are added to the set of R-successors of d.

Now, J' is defined by

$$\begin{aligned}
dom(J') &:= \bigcup_{i \geq 0} dom(J_i) \\
R^{J'} &:= \bigcup_{i \geq 0} R^{J_i} \\
P^{J'} &:= \{d \in dom(J') \mid \varepsilon \in L_{d,J'}(P)\}.
\end{aligned}$$

The extended canonical model I' of $\mathcal{A}$ is induced by J'. ◇

We call elements in $dom(J')\setminus dom(J)$ *extension elements*. An individual $e \in dom(J')$ is *generated in* J_k if $k = 0$ and $e \in dom(J_k)$ or if $e \in dom(J_k) \setminus dom(J_{k-1})$ for $k \geq 1$. Note that there is exactly one $k \geq 0$ such that e is generated in J_k. If $i < j$, then we say that the individuals generated in J_j are *generated later* than the individuals in J_i.

In order to show that I' is a model of $\mathcal{A}$ the following condition is necessary:

> The knowledge base $\mathcal{A}$ is consistent; $W \notin E_a$ and if $r > 0$, then also $WR \notin E_a$; $c^{\leq R}_{a,W} \geq r$; if $W \in S \cdot N^*_R$, then $|S_{\mathcal{A}}(a)| < c^{\leq S}_{a,\varepsilon}$. (1)

We summarize the main properties of J' and I' in

Lemma 37.
Using the denotation in the above definition it holds that:

1. Extension elements only have extension elements as successors which in addition are generated later. (These successors form a tree.)

2. For all $d \in dom(J)$ it holds that

$$\begin{aligned} c^{\geq S}_{d,J} &= c^{\geq S}_{d,J'} = c^{\geq S}_{d,J_i} \text{ for all } i \geq 0, \\ c^{\leq S}_{d,J} &= c^{\leq S}_{d,J'} = c^{\leq S}_{d,J_i} \text{ for all } i \geq 0. \end{aligned}$$

3. All paths in J' which lead from an individual in $\mathcal{A}$ to an extension element (or to d_1) have as suffix a path from $a^{I'}$ to this extension element (or to d_1). Furthermore, if this suffix is of minimal length it has as prefix the path $a^{I'}U^{I'}d_1$.

4. The direct successors of d_1 are only individuals in dom(J) and d_2 or $f_1, \ldots, f_k$ (if $W = U$).

5. For all $d \in dom(J_i)$, $i \geq 0$, it holds that

$$\begin{aligned} c^{\geq S}_{d,J_i} &= c^{\geq S}_{d,J_j} \text{ for all } j \geq i, \\ c^{\geq S}_{d,J_i} &= c^{\geq S}_{d,J'}, \\ c^{\leq S}_{d,J_i} &= c^{\leq S}_{d,J_j} \text{ for all } j \geq i, \\ c^{\leq S}_{d,J_i} &= c^{\leq S}_{d,J'}. \end{aligned}$$

6. For an individual generated in J_i, $i \geq 0$, all later generated direct successors are generated in J_{i+1}. In particular, at most in one iteration step in the construction of J' direct successors of an individual are generated.

7. Let $d \in dom(J')$, $S \in N_R$. Then, if $|S^{J_0}(d)| \leq c^{\geq S}_{d,J'}$, then $|S^{J'}(d)| = c^{\geq S}_{d,J'}$.

8. If $d \in (dom(J') \setminus dom(J)) \cup \{d_1\}$, then for all $S \in N_R$ there is exactly one $X \in N_R^*$ such that $a^{J'} X^{J'} d$ where $a^{J'}$ only occurs as initial node in this path. Furthermore, for X it is $c^{\leq S}_{d,J'} \geq c^{\leq S}_{a,X}$ and $c^{\geq S}_{d,J'} \leq c^{\geq S}_{a,X}$.

9. There is an individual $d \in dom(I')$ such that $a^{I'} W^{I'} d$ and $|R^{I'}(d)| \geq r$.

10. If the conditions in (1) are satisfied, then the extended canonical model I' is a model of $\mathcal{A}$.

Proof. 1.: This is an easy consequence of the construction of J'

2.: Since J_i, $i \geq 0$, and J' are extensions of J, every path from an individual in $\mathcal{A}$ to d in J is also a path in J_i, $i \geq 0$, and J'. Thus, $c^{\geq S}_{d,J} \leq c^{\geq S}_{d,J'}$ and $c^{\geq S}_{d,J} \leq c^{\geq S}_{d,J_i}$ for all $i \geq 0$. On the other hand, because of 1., every path from an individual in $\mathcal{A}$ to d in J' and J_i is also a path in J. Consequently, $c^{\geq S}_{d,J} \geq c^{\geq S}_{d,J'}$ and $c^{\geq S}_{d,J} \geq c^{\leq S}_{d,J_i}$ for all $i \geq 0$. For at-most restrictions the claim can be shown analogously.

3.: By Lemma 26, 5. for all $d \in dom(J)$ and $S \in N_R$ we know $|S^J(d)| \geq c^{\geq S}_{d,J}$. Thus, using 2., no direct successors for d are generated, and consequently, all extension elements are successors of d_1, and every path from an element in $dom(J)$ to an extension element contains d_1. Lemma 26, 3. implies that every path in J' from an element in $\mathcal{A}$ to d_1, which by 1. is also a path in J, contains $a^{J'}$. This shows the first part of 3. Consequently, a path in J' from $a^{J'}$ to an extension element contains d_1. Furthermore, by 1. the path from $a^{J'}$ to d_1 is also a path in J. If $a^{J'}$ occurs only as initial node, then this path is a new role chain and by Lemma 26, 3. is uniquely determined with label U (see the construction of J_0 in Definition 36). By construction of J' the path in J' from d_1 to the extension element is uniquely determined as well. Thus the path from $a^{J'}$ to the extension element or to d_1 is uniquely determined and has $a^{J'} U^{J'} d_1$ as prefix.

4.: From 2. we can deduce that for d_1 no direct successors are generated since $|S^J(d_1)| \geq c^{\geq S}_{d_1,J} = c^{\geq S}_{d_1,J_i}$ for all $i \geq 0$. Thus, the only direct successors of d_1 which are extension elements are those defined in J_0, namely, d_2 or $f_1, \ldots, f_k$.

5.: Let b be an individual in $\mathcal{A}$, $V \in N_R^*$, $i \leq j$, and $d \in dom(J_i)$. Then,

Claim: $b^{J_i} V^{J_i} d$ iff $b^{J_j} V^{J_j} d$.

Proof of the Claim: For $i = j$ there is nothing to show. Assume that $i < j$. Since J_j is an extension of J_i, the only-if direction of the claim is trivial. All elements generated in J_j are, by definition, generated later than those in J_i. Thus, by 1. we can conclude that $b^{J_j} V^{J_j} d$ contains no elements generated in J_j. This shows the if direction of the claim.

Using the claim, statement 5. can easily be shown.

6.: Let d be an individual generated in J_i, $i \geq 0$. According to the construction we know $|S^{J_{i+1}}(d)| \geq c^{\geq S}_{d,J_i}$. Now 5. implies that no direct successors for d are generated in J_j, $j > i + 1$.

7.: Let $|S^{J_0}(d)| \leq c^{\geq S}_{d,J'}$ and $d \in dom(J')$. Then, there is a number $i \geq 0$ such that d is generated in J_i. Thus, $|S^{J_0}(d)| = |S^{J_i}(d)|$. Using 5., $|S^{J_0}(d)| \leq c^{\geq S}_{d,J'}$ implies $|S^{J_i}(d)| \leq c^{\geq S}_{d,J_i}$. By construction it follows $|S^{J_{i+1}}(d)| = c^{\geq S}_{d,J_i}$. By 6. we know $|S^{J_{i+1}}(d)| = |S^{J'}(d)|$ and by 5. we have $c^{\geq S}_{d,J_i} = c^{\geq S}_{d,J'}$. Consequently, $|S^{J'}(d)| = c^{\geq S}_{d,J'}$.

8.: By the construction of J' it is easy to see that for d there is a path from an individual in $\mathcal{A}$ to d. By 3. this path contains $a^{I'}$. This shows that there is $X \in N_R^*$ with $a^{J'}X^{J'}d$. We can assume that X is of minimal length, i.e., $a^{J'}$ only occurs as initial node of the path $a^{J'}X^{J'}d$. Let $m := c^{\leq S}_{d,J'}$. If $m = \infty$, we know $m \geq c^{\leq S}_{a,X}$. Assume $m < \infty$. Then by definition of $c^{\leq S}_{d,J'}$ there is a concept assertion $F(f) \in \mathcal{A}$ and a word $Y \in N_R^*$ such that $Y \in L_F((\leq m\ R))$ and $f^{J'}Y^{J'}d$. By 3. we know that $f^{J'}Y^{J'}d$ contains $a^{J'}$ and $a^{J'}X^{J'}d$ is a suffix of $f^{J'}Y^{J'}d$. Consequently, there is a $Y' \in N_R^*$ such that $Y = Y'X$ and $f^{J'}Y'^{J'}a^{J'}X^{J'}d$. Using 1. and Lemma 26, 1. the path $f^{J'}Y'^{J'}a^{J'}$ is a path in $\mathcal{A}$. Hence, we know $m \geq c^{\leq S}_{a,X}$. Analogously, we can conclude $c^{\geq S}_{d,J'} \leq c^{\geq S}_{a,X}$.

9.: This claim is an immediate consequence of the construction of J_0.

10.: By definition of I' the role assertions in $\mathcal{A}$ are satisfied. Now let $B(b) \in \mathcal{A}$. We have to show that $b^{I'} \in B^{I'}$ provided that the conditions in (1) hold. We use Theorem 22 in [10] and distinguish the following cases:

(P1): Let P be a primitive concept, $Y \in L_B(P)$, $d \in dom(J')$ where $b^{J'}Y^{J'}d$. By definition of $L_{d,J'}(P)$ it follows $\varepsilon \in L_{d,J'}(P)$, and hence, $d \in P^{I'}$.

(P2): Let $(\geq m\ S)$ be a at-least restriction, $Y \in L_B((\geq m\ S))$, $d \in dom(J')$ where $b^{J'}Y^{J'}d$. This implies $c^{\geq S}_{d,J'} \geq m$. By 7. we can conclude $|S^{J'}(d)| \geq c^{\geq S}_{d,J'}$, and thus, $d \in (\geq m\ S)^{I'}$.

(P3): Let $(\leq m\ S)$ be a at-most restriction, $Y \in L_B((\leq m\ S))$, $d \in dom(J')$ where $b^{J'}Y^{J'}d$. Assume that $d \notin (\leq m\ S)^{I'}$, i.e., $|S^{I'}(d)| > m$. We distinguish the following cases:

(i) Assume $d \in dom(J) \setminus \{d_1\}$. As a consequence of 1. the path $b^{I'}Y^{I'}d$ is also a path in I. Furthermore, 2. implies $S^{I'}(d) = S^{I}(d)$. But then we have $b^I \notin B^I$, and thus, I is no model of $\mathcal{A}$, which is a contradiction to the fact that $\mathcal{A}$ is consistent and the statement in Lemma 26, 6., namely, I is a model of $\mathcal{A}$.

(ii) Assume $d = d_1$ and $n \geq 1$. (For the case $n = 0$ see the case $d = d_{n+1}$.) As shown in 4. for d_1 no direct successors are generated. Thus, if $S \neq R_1$, then we have $S^{I'}(d) = S^{I}(d)$, and consequently, $d \notin (\leq m\ S)^I$. Furthermore, 1. implies that the path $b^{I'}Y^{I'}d$ is also a path in I. Again, this is a contradiction to the fact that I is a model of $\mathcal{A}$. Now, assume $S = R_1$. We distinguish two cases:

  (I) The word U in Definition 36 is not the empty word. Then $|S^I(d)| = 0$, because U was chosen maximal. Further, by 2. we know $c^{\geq S}_{d_1,J} = c^{\geq S}_{d_1,J_i}$ for all $i \geq 0$. Thus, no S-successors are generated for d_1, and by construction we can conclude $|S^{J'}(d_1)| = 1$. But then $m = 0$. By 3. the path $b^{J'}Y^{J'}d_1$ contains $a^{J'}$, say $b^{J'}Y'^{J'}Y''^{J'}d_1$ where $Y = Y'Y''$, $b^{J'}Y'^{J'}a^{J'}$ is a chain in $\mathcal{A}$. We can assume Y'' to be of minimal length. But then, the path $a^{J'}Y''^{J'}d_1$ only contains $a^{J'}$ as initial node. By 3. this implies

$Y'' = U$. Consequently, $c^{\leq S}_{a,U} = 0$. Since US is a prefix of W this is a contradiction to $W \notin E_a$.

(II) Assume $U = \varepsilon$. Then by construction it is $d_1 = a^{J'}$. By definition of U we know that for d no S-successors are generated in J, i.e., d has no S-successors in J which are not individuals in $\mathcal{A}$. This means $|S_{\mathcal{A}}(a)| \geq c^{\geq S}_{a,J}$. By 2. we have $c^{\geq S}_{a,J} = c^{\geq S}_{a,J_i}$ for every $i \geq 0$. This implies that even in J' no S-successors are generated for d. Then, we can conclude $|S^{J_0}(d)| = |S^{J'}(d)| = |S_{\mathcal{A}}(a)| + 1$. By the assumption we know $|S_{\mathcal{A}}(a)| < c^{\leq S}_{a,\varepsilon}$. Furthermore, by 8. we know $c^{\leq S}_{a,\varepsilon} \leq c^{\leq S}_{a,J'}$. This shows $|S^{J'}(d)| \leq c^{\leq S}_{a,J'}$. Additionally, using $b^{J'}Y^{J'}d$ (by 3. this path contains $a^{J'}$), $Y \in L_B((\leq m\ S))$, and $B(b) \in \mathcal{A}$ it follows $c^{\leq S}_{a,J'} \leq m$. But then, $|S^{J'}(d)| \leq m$. This is a contradiction to the assumption.

(iii) Assume $d \in \{d_2, \ldots, d_n\}$. We distinguish two cases:

(I) Assume $S \neq R_i$ and $d = d_i$ where $2 \leq i \leq n$. By construction we know $|S^{J_0}(d)| = 0$. Now 7. implies $|S^{J'}(d)| = c^{\geq S}_{d,J'} > m$. On the other hand, as in (ii), (II) we can conclude $c^{\leq S}_{d,J'} \leq m$. The path $b^{J'}Y^{J'}d$ contains $a^{J'}$ (3.). But then, by 8. there is a word X such that $a^{J'}X^{J'}d$ where $a^{J'}$ occurs only as initial node in this path, $c^{\leq S}_{a,X} \leq c^{\leq S}_{d,J'}$ and $c^{\geq S}_{a,X} \geq c^{\geq S}_{d,J'}$. Consequently, we have $c^{\leq S}_{a,X} \leq m < c^{\geq S}_{a,X}$. Furthermore, since X is uniquely determined, X is a prefix of W. Thus, $W \in E_a$, in contradiction to the assumption.

(II) Now, let $S = R_i$ and $d = d_i$ where $2 \leq i \leq n$. If $|S^{J_0}(d)| \leq c^{\geq S}_{d,J'}$, then we can proceed as in (iii), (I). Now, assume $|S^{J_0}(d)| > c^{\geq S}_{d,J'}$. Since $|S^{J_0}(d)| = 1$, we know $c^{\geq S}_{d,J'} = 0$. Because of $c^{\geq S}_{d,J'} = 0$, and thus, $c^{\geq S}_{d,J_i} = 0$ for all $i \geq 0$, no direct S-successors of d are generated. Hence, $|S^{J'}(d)| = 1$. Using $|S^{J'}(d)| > m$ it follows $m = 0$. By 3. the path $b^{J'}Y^{J'}d_i$ contains $a^{J'}$, say $b^{J'}Y'^{J'}Y''^{J'}d_i$ where $Y = Y'Y''$, $b^{J'}Y'^{J'}a^{J'}$ is a chain in A. Furthermore, by 3. we know $Y'' = UR_1 \cdots R_{i-1}$. Consequently, $c^{\leq R_i}_{a,UR_1\cdots R_{i-1}} = 0$. Since $UR_1 \cdots R_i$ is a prefix of W this is a contradiction to $W \notin E_a$.

(iv) Assume $d = d_{n+1}$.

(I) We assume $S \neq R$. Then, we can show a contradiction as in (iii), (I).

(II) We assume $S = R$. We consider two cases.

(a) Let $|R^{J_0}(d)| \leq c^{\geq R}_{d,J'}$. Then, by 7. we have $|R^{J'}(d)| = c^{\geq R}_{d,J'} > m$. On the other hand, as in (ii), (II) we can conclude $c^{\leq S}_{d,J'} \leq m$. Analogously to (iii), (I), this leads to an contradiction.

(b) Let $|R^{J_0}(d)| > c^{\geq R}_{d,J'}$. Since $c^{\geq R}_{d,J'} = c^{\geq R}_{d,J_i}$ for all $i \geq 0$ (5.), no R-successors are generated for d. Hence, $|R^{J'}(d)| = |R^{J_0}(d)|$. We distinguish two cases.

(A) Assume $|R^{J_0}(d)| = r > m$. According to 8., we know $c^{\leq R}_{d,J'} \geq c^{\leq R}_{a,W}$. Moreover, as in (ii),(II) we can conclude $c^{\leq R}_{d,J'} \leq m$. On the other hand, by assumption we have $c^{\leq R}_{a,W} \geq r$. Thus, $r > m \geq c^{\leq R}_{d,J'} \geq c^{\leq R}_{a,W} \geq r$ which is a contradiction.

(B) Suppose $|R^{J_0}(d)| > r$. By construction this implies $R^J(d) = R^{J_0}(d)$, $d = d_1 \in dom(J)$, and $a^J U^J d$. Since I is a model of $\mathcal{A}$, we know by Lemma 16 that (*) $c^{\geq R}_{a,U} \leq |R^J(d)| \leq c^{\leq R}_{a,U}$. As above $c^{\leq R}_{d,J'} \leq m$. By 8. this implies $c^{\leq R}_{a,U} \leq m$, which is a contradiction to (*) and $|R^J(d)| > m$.

(v) Assume $d \in \{f_1, \ldots, f_k\}$. By construction we know $|S^{J_0}(d)| = 0$. Thus, 7. implies $|S^{J'}(d)| = c^{\geq S}_{d,J'} (> m)$. As above we have $c^{\leq S}_{d,J'} \leq m$. According to 8. we know $c^{\leq S}_{a,WR} \leq c^{\leq S}_{a,J'}$ and $c^{\geq S}_{a,WR} \geq c^{\geq S}_{a,J'}$. Hence, $c^{\leq S}_{a,WR} < c^{\geq S}_{a,WR}$. Consequently, $WR \in E_a$, which is a contradiction to the assumption.

(vi) Suppose $d \in dom(J') \setminus (dom(J) \cup \{d_2, \ldots, d_{n+1}, f_1, \ldots, f_k\})$. By construction we know $|S^{J_0}| = 0$. Thus, by 7. $|S^{J'}(d)| = c^{\geq S}_{d,J'} (> m)$. As above we have $c^{\leq S}_{d,J'} \leq m$. According to 8. there is a word Y' such that $c^{\leq S}_{a,Y'} \leq c^{\leq S}_{d,J'}$ and $c^{\geq S}_{a,Y'} \geq c^{\geq S}_{d,J'}$. Consequently, $c^{\geq S}_{a,Y'} > c^{\leq S}_{a,Y'}$. By construction of J' the word Y' is of the form $Y' = V'V''$ where V' is a maximal prefix of W (in case of $r = 0$) or WR (in case of $r > 0$) and $a^{J'} V'^{J'} g V''^{J'} d$ with $g \in \{d_2, \ldots, d_{n+1}, f_1, \ldots, f_k\}$ (The element d_1 is not contained in this set because of 4. Let $V'' = Q_1 \cdots Q_s$, $s \geq 1$, and $e_1, \ldots, e_{s-1}$ elements with $e_0 Q_1^{J'} e_1 \cdots Q_s^{J'} e_s$ where $e_0 := g$ and $e_s := d$. By 6. the element e_j, $0 \leq j \leq s$, is generated in J_j. Thus, for $0 \leq j \leq s-1$ we have $(c^{\geq Q_{j+1}}_{e_j,J_j} - |Q_{j+1}(e_j)|) > 0$. In particular, it follows $c^{\geq Q_{j+1}}_{e_j,J_j} \geq 1$, and by 8. this implies $c^{\geq Q_{j+1}}_{a,V'Q_1\cdots Q_j} \geq 1$ for all $0 \leq j \leq s-1$. Consequently, $V'V''$ is required by a starting from V' and $c^{\geq S}_{a,V'V''} > c^{\leq S}_{a,V'V''}$. This is a contradiction to the fact that $W \notin E_a$ ($r = 0$) or $WR \notin E_a$ ($r > 0$).

□

Now, we can continue the proof of the only-if direction in Theorem 35. We consider three cases.

1. Assume $W \notin V_b(P)$. Then, there is a $c \in I_{\mathcal{A}}$, V, V' such that $W = VV'$, $V \in L(b,c)$, and $V' \in N_c \cap \overline{L_c(P) \cup E_c}$. But then, by Lemma 37 $I' = I(c, V')$ is a model of $\mathcal{A}$ since the conditions in (1) are satisfied. Furthermore, there is an element $d (= d_{n+1})$ such that $b^{I'} V^{I'} c^{I'} V'^{I'} d$. Assume $d \in P^{I'}$, thus, $\varepsilon \in L_{d,J'}(P)$. Hence, there is a concept assertion $F(f) \in \mathcal{A}$ and a word $U \in N_R^*$ such that $U \in L_F(P)$ and $f^{I'} U^{I'} d$. Since $d = d_{n+1}$, the path $f^{I'} U^{I'} d$ contains $c^{I'}$ (Lemma 26, 3. and Lemma 37, 3.). This means, there are words U', U'' such that $U = U'U''$ and $f^{I'} U'^{J'} c^{I'} U''^{I'} d$ where $c^{I'}$ occurs in $c^{I'} U''^{I'} d$ only as initial node. The path $fU'c$ is a role chain in $\mathcal{A}$ (Lemma 26, 1.). Thus, we have $U'' \in L_c(P)$. Since the path from $c^{I'}$ to d is uniquely determined if $c^{I'}$ only occurs as initial node, it

follows $U'' = V'$. But then, $V' \in L_c(P)$, which is a contradiction. This implies $d \notin P^{I'}$, and thus, $b^{I'} \notin (\forall W.P)^{I'}$.

2. Assume $W \notin V_b(\geq l\ R)$, $l > 0$. Then, there is a $c \in I_{\mathcal{A}}$, V, V' such that $W = VV'$, $V \in L(b,c)$, and $V' \in N_c(\geq l\ R) \cap \overline{\bigcup_{r \geq l} L_c(\geq rR) \cup E_c}$. But then, by Lemma 37 $I' = I(c, V')$ is a model of $\mathcal{A}$ since the conditions in (1) are satisfied. Furthermore, there is an element $d (= d_{n+1})$ such that $b^{I'} V^{I'} c^{I'} V'^{I'} d$ where $c^{I'}$ occurs only as initial node in $c^{I'} V'^{I'} d$. Assume $d \in (\geq l\ R)^{I'}$, i.e., $|R^{I'}(d)| \geq l$. We consider two cases.

   (a) Suppose $|R_{\mathcal{A}}(d)| \geq l$. Then, $d \in I_{\mathcal{A}}$ (Lemma 26, 1. and Lemma 37, 1.). Furthermore, by construction of I' and since $d = d_{n+1}$ it follows $d = c$ and $V' = \varepsilon$. But then, $V' = \varepsilon \in N_c(\geq l\ R)$ yields $|R_{\mathcal{A}}(d)| < l$, which is a contradiction to the assumption.

   (b) Assume $|R^{J'}(d)| \geq l$ and $|R_{\mathcal{A}}(d)| < l$. If $d \in dom(J)$, then no R-successors are generated for d in J' (proof of Lemma 37, 3.), i.e. $R^{J'}(d) = R^J(d)$. On the other hand, R-successors are generated for d in J because of $|R^{J'}| \geq l$ and $|R_{\mathcal{A}}(d)| < l$. This shows, $|R^J(d)| = c^{\geq R}_{d,J}$ (Lemma 26, 5.). Using Lemma 37, 2. this yields $|R^{J'}(d)| = c^{\geq R}_{d,J'}$. If d is an extension element in J', then we have $|R^{J_0}(d)| = 0$ since $d = d_{n+1}$. Again, we have $|R^{J'}(d)| = c^{\geq R}_{d,J'}$ (Lemma 37, 7.). According to Lemma 37, 8. we have $c^{\geq R}_{c,V'} \geq c^{\geq R}_{d,J'}$. Hence, $c^{\geq R}_{c,V'} \geq l$. Consequently, there is an $r \geq l$ where $V' \in L_c(\geq rR)$, which is a contradiction.

   Thus, we have shown $|R^{I'}(d)| < l$, and $b^{I'} \notin (\forall W.(\geq l\ R))^{I'}$.

3. Assume $W \notin V_b(\leq l\ R)$. Then, there is a $c \in I_{\mathcal{A}}$, V, V' such that $W = VV'$, $V \in L(b,c)$, and $V' \in N_c \cap \overline{\bigcup_{r \leq l} L_c(\leq rR) \cup (E_c \cdot R^{-1})}$. But then, by Lemma 37 $I' = I(c, V', R, l+1)$ is a model of $\mathcal{A}$ since the conditions in (1) are satisfied. Furthermore, there is an element $d (= d_{n+1})$ such that $b^{I'} V^{I'} c^{I'} V'^{I'} d$ and $|R^{I'}(d)| \geq l+1$ (Lemma 37, 9.). This shows $d \notin (\leq l\ R)^{I'}$. Hence, $b^{I'} \notin (\forall W.(\leq l\ R))^{I'}$.

This completes the proof of Theorem 35.

In the next section we use value-restriction sets to decide instance and to construct the most specific concept of an individual. For that purpose, we have to show that value-restriction sets are regular. Obviously, it is sufficient to show that the sets $L_c(P)$, N_c, and E_c are regular.

Proposition 38.
Predecessor restriction sets are regular.

Proof. It is easy to see that predecessor restriction sets for primitive concepts or number-restrictions P can be described as follows:

$$L_a(P) \quad = \quad \bigcup_{\substack{F(f) \in \mathcal{A} \\ A \in \mathcal{C} \\ L_f(a) \cap L_F(A) \neq \emptyset}} L_A(P)$$

where $\mathcal{C}$ is defined as in Section 4 (page 15). Thus, predecessor restriction sets are finite union of regular languages. Thus, they are regular. An automaton accepting $L_a(P)$ can be constructed as follows: Let T be a TBox containing equivalent defined concepts for all $\mathcal{FLN}$-concept descriptions occurring in $\mathcal{A}$. Then the corresponding semi-automaton $\mathcal{A}_T$ accepts the predecessor restriction set $L_a(P)$ if P is the finite state of $\mathcal{A}_T$ and the set of concepts $A \in \mathcal{C}$ such that there is an $F(f) \in \mathcal{A}$ with $L_f(a) \cap L_F(A) \neq \emptyset$ is the initial set of $\mathcal{A}_T$. This automaton is linear in the size of $\mathcal{A}$ and can be constructed using polynomial space. □

Obviously, the sets N_c and $N_c(\geq n\ R)$ are regular sets. In order to show that finite automata accepting these languages can be constructed using polynomial space we only have to show that number conditions can be computed in space polynomial in the size of $\mathcal{A}$.

Proposition 39.
Number conditions $c_{c,V}^{\geq R}$ and $c_{c,V}^{\leq R}$ are computable using polynomial space in the size of $\mathcal{A}$ and V.

Proof. The finite automaton accepting $L_c(\geq nR)$ can be constructed using polynomial space (see the proof of Proposition 38). Thus, testing $V \in L_c(\geq nR)$ can be done in time polynomial in the size of V and $\mathcal{A}$. Consequently, $max\{n \,|\, V \in L_c(\geq nR)\}$ is computable using polynomial space. For $c_{c,V}^{\leq R}$ the claim can be verified analogously. □

Finally, we have to show that E_c is regular. For this purpose, we construct a finite automaton accepting E_c using the following notion.

Definition 40 (reaching p-exclusion sets).
The word W *reaches a p-exclusion set starting from* c if there is a prefix V of W such that $next_\varepsilon(q_c, V) \in \mathcal{E}$. ◇

Now we can characterize E_c as follows:

Lemma 41 (characterizing E_c).

$E_c = \{W \in N_R^* \,|\,$ starting from c a p-exclusion set is reachable by W, or there is a prefix VR of W where $V \in N_R^*$, $R \in N_R$ such that $(\leq 0\ R) \in next_\varepsilon(q_c, V)\}$

Proof. "⊆" Let $W \in E_c$. We distinguish two cases:

1. There exist $V, V' \in N_R^*$, $R \in N_R$ such that V is a prefix of W, VV' is required by c starting from V, and $c_{c,VV'}^{\geq R} > c_{c,VV'}^{\leq R}$. Similar to 2. in the proof of Proposition 29 one can show that for $F_i := next_\varepsilon(q_c, VR_1 \cdots R_i)$, $0 \leq i \leq n$ it holds that $(\geq n_i\ R_i) \in F_{i-1}$ for all $1 \leq i < n$ where $n_i \geq 1$. Furthermore, for $l := c_{c,VV'}^{\geq R}$ and $r := c_{c,VV'}^{\leq R}$ we can conclude $(\geq l\ R), (\leq r\ R) \in F_n$. Thus, F_0 is a p-exclusion set that is reachable by V starting from c. This means, W reaches a p-exclusion set starting from c.

2. There is a prefix VR of W, $V \in N_R^*$, $R \in N_R$, such that $c_{c,V}^{\leq R} = 0$. As in 2. in the proof of Proposition 29 it can be shown that $(\leq 0\ R) \in next_\varepsilon(q_c, VR)$.

"$\supseteq$" Let W be a word in the right-hand side of the equation in the claim. We distinguish two cases:

3. There is a prefix V of W such that $next_\varepsilon(q_c, V) \in \mathcal{E}$. By definition of $\mathcal{E}$ there is a word $V' = R_1 \cdots R_n \in N_R^*$, conflicting number-restrictions $(\geq l\ R)$ and $(\leq r\ R)$, $l > r$, and for all $1 \leq i \leq n$ there are integers $m_i \geq 1$ such that for $F_0 := next_\varepsilon(q_c, V)$ and $F_i := next_\varepsilon(F_0, R_1 \cdots R_i)$, $1 \leq i \leq n$, it is $(\geq m_i\ R_i) \in F_{i-1}$ for all $1 \leq i \leq n$ and $(\geq l\ R), (\leq r\ R) \in F_n$. Consequently, there are $A_1, \ldots, A_n$ as well as A^l, A^r in q_c such that $VR_1 \cdots R_{i-1} \in L_{A_i}((\geq m_i\ R_i))$ for all $1 \leq i \leq n$ as well as $VV' \in L_{A^l}((\geq l\ R))$ and $VV' \in L_{A^r}((\leq r\ R))$. By definition of q_c there are atomic concepts F_i, individuals f_i, and words W_i for all $1 \leq i \leq n$ as well as atomic concepts F^l, F^r, individuals f^l, f^r, and words W^l, W^r such that $F_i(f_i) \in \mathcal{A}$, $f_i W_i c$ role chain in $\mathcal{A}$, and $W_i \in L_{F_i}(A_i)$ as well as $F^l(f^l) \in \mathcal{A}$, $f^l W^l c$ role chain in $\mathcal{A}$, $W^l \in L_{F^l}(A^l)$, and $F^r(f^r) \in \mathcal{A}$, $f^r W^r c$ role chain in $\mathcal{A}$, $W^r \in L_{F^r}(A^r)$. This means, $W_i V R_1 \cdots R_{i-1} \in L_{F_i}((\geq m_i\ R_i))$ for all $1 \leq i \leq n$ as well as $W^l V V' \in L_{F^l}((\geq l\ R))$ and $W^r V V' \in L_{F^r}((\leq r\ R))$. Consequently, VV' is required by c starting from V and $c^{\geq R}_{a,VV'} \geq l > r \geq c^{\leq R}_{c,VV'}$. Thus, $W \in E_c$.

4. There is a prefix VR of W where $V \in N_R^*$, $R \in N_R$ such that $(\leq 0\ R) \in next_\varepsilon(q_c, V)$. Thus, there is a $A \in q_c$ such that $V \in L_A((\leq 0\ R))$. As in 3. by definition of q_c it can be shown $c^{\leq R}_{c,V} \leq 0$, hence, $c^{\leq R}_{c,V} = 0$. This means, $W \in E_c$.

□

Using the description of E_c shown in the above Lemma we construct a finite automaton accepting E_c. Therefore, we need the notion "powerset automaton" which is well-known from automata theory. We introduce this notion for the semi-automaton $\mathcal{A}_T$ and an individual c.

Definition 42 (Powerset automaton of individuals).
Let $\mathcal{A}_T = (N_R, Q, E)$ denote the semi-automaton corresponding to the $\mathcal{FLN}$-terminology T (where T is defined as in the proof of Proposition 38). Let c be an individual in $\mathcal{A}$. Then the *powerset automaton* $\mathcal{P}(\mathcal{A}_T, c)$ *of* $\mathcal{A}_T$ *and* c is defined by $\mathcal{P}(\mathcal{A}_T, c) := (N_R, \widehat{Q}, \widehat{q}, \widehat{\delta})$ where

- $\widehat{Q} := \{G \subseteq Q \mid next_\varepsilon(q_c, W) = G \text{ for a } W \in N_R^*\}$ (set of states), and
- $\widehat{q} := q_c$ (initial state), and
- $\widehat{\delta}(I, R) := next_\varepsilon(I, R) \in \widehat{Q}$ for $I \in \widehat{Q}$, and $R \in N_R$ (set of transitions).[5]

◇

By Lemma 41 it can easily be verified that the finite automaton $\mathcal{B}_c$, defined in the following definition, accepts the language E_c:

[5]Note that we do not define final states.

Definition 43 (a finite automaton for E_c).
The automaton $\mathcal{B}_c$ is defined by the powerset automaton $\mathcal{P}(\mathcal{A}_T, c)$ of $\mathcal{A}_T$ and c where a new state q is added, which is the final state of $\mathcal{B}_c$. Furthermore, for every p-exclusion set $F \subseteq Q$ in $\mathcal{B}_c$ we add a transition (F, ε, q) and for every state $F \subseteq Q$ in $\mathcal{B}_c$ and at-most restriction $(\leq 0\ R)$ in T such that $(\leq 0\ R) \in F$ we add the transition (F, R, q). Finally, we add (q, R, q) for every $R \in N_R$. ◇

By this finite automaton that accepts E_c we have shown that E_c is regular. Note that $\mathcal{B}_c$ is exponential in the size of $\mathcal{A}$ and can be constructed in time exponential in the size of $\mathcal{A}$. Furthermore, it is not hard to see that a finite automaton for $E_c R^{-1}$ can be computed in time exponential in the size of $\mathcal{A}$ as well. To sum up, we have shown

Proposition 44.
Value-restriction sets are regular and the corresponding finite automata are effectively computable. □

5.2 Deciding instance

Using Theorem 35 it is not hard to characterize instance.

Theorem 45 (instance).
Let C be a cyclic $\mathcal{FLN}$-concept description and b be an individual name in $\mathcal{A}$, then $b \in_{\mathcal{A}} C$ iff for all primitive concepts and number restrictions P in $\mathcal{A}$ it holds that $L_C(P) \subseteq V_b(P)$.

Proof. "⇐" Assume that $L_C(P) \subseteq V_b(P)$ holds for all primitive concepts and number restrictions P. By the characterization of the gfp-semantics [10] we have to show that $W \in L_C(P)$ implies $b \in_{\mathcal{A}} \forall W.P$. Let $W \in L_C(P)$. By the assumption we know that $W \in V_b(P)$. By Theorem 35 this yields $b \in_{\mathcal{A}} \forall W.P$.

"⇒" Assume that there is a primitive concept or a number restriction P and a word $W \in L_C(P) \setminus V_b(P)$. By Theorem 35 this implies $b \notin_{\mathcal{A}} \forall W.P$. Using the characterization of the gfp-semantics [11], $W \in L_C(P)$ implies $b \notin_{\mathcal{A}} C$. □

In order to decide instance we only have to test the inclusions stated in Theorem 45. By Proposition 44 we know that the languages in the inclusions are regular and that the corresponding finite automata are effectively computable. Thus, instance is decidable. In [2], it has been shown that instance w.r.t. ABoxes disallowing number-restrictions is PSPACE-complete (by reducing subsumption of atomic concepts defined in cyclic terminologies to instance). Thus, we can conclude

Corollary 46 (complexity of instance).
Deciding instance w.r.t. Aboxes with cyclic $\mathcal{FLN}$-concept descriptions is PSPACE-hard. □

5.3 Computing the msc

The msc of an individual must contain the maximal set of value-restrictions which are satisfied by the individual. Since we already have defined such sets, namely, the value-restriction sets, it is easy to prove the following theorem using Theorem 45 and Proposition 11.

Theorem 47 (most specific concept).
Let $\mathcal{A}$ be consistent, C be a cyclic $\mathcal{FLN}$-concept description, and b an individual occurring in $\mathcal{A}$. Then C is the msc of b in $\mathcal{A}$ if for all primitive concepts and number restrictions P we have $L_C(P) = V_b(P)$.[6]

Proof. If for all primitive concepts and number restrictions P we have $L_C(P) = V_b(P)$, then by Theorem 45 it follows $b \in_{\mathcal{A}} C$. Furthermore, if $b \in_{\mathcal{A}} D$, then $L_D(P) \subseteq V_b(P)$, thus, $L_D(P) \subseteq L_C(P)$. Now, $L_C(P) \subseteq V_C(P)$ implies $L_D(P) \subseteq V_C(P)$, thus, by Proposition 11 we have $C \sqsubseteq D$. □

As an immediate consequence of this Theorem and Proposition 44 we have

Corollary 48.
Most specific concepts for individuals defined in Aboxes with cyclic $\mathcal{FLN}$-concept descriptions are effectively computable. □

6 Related and future work

An important topic for future work is to determine the exact worst-case complexities for computing the lcs and the msc, and for deciding the instance problem for $\mathcal{FLN}$-ABoxes with cyclic concept descriptions. Our algorithm for computing the lcs of two cyclic $\mathcal{FLN}$-concept descriptions is exponential, and we conjecture that this complexity cannot be avoided, i.e., there is no polynomial algorithm for computing the lcs in this case. One point supporting this conjecture is that subsumption for cyclic $\mathcal{FLN}$-concept descriptions is already PSPACE-complete (see [11]). It is, however, not clear how to reduce the subsumption problem (in polynomial time) to the problem of computing the lcs. In fact, if $C \sqsubseteq D$, then the lcs of C and D is equivalent to D, but testing for this equivalence may be as hard as testing for subsumption.

A naive analysis of the algorithms for deciding the instance problem and for computing the msc derived from our characterization of value-restriction sets would yield a triply exponential upper bound. In fact, the first exponential step is due to the fact that an automaton for E_a may already be exponential in the size of the input. The other two exponential steps are due to the two complements occurring in the characterization of the value-restriction sets.[7] We conjecture, however, that the instance problem can be decided in PSPACE, and that the msc can be computed in exponential time.

[6] If $\mathcal{A}$ is inconsistent, then $C = \bot$ is the most specific concept of b w.r.t. $\mathcal{A}$.

[7] since computing the complement of a regular language requires a powerset construction.

To the best of our knowledge, all the existing work on computing the lcs of description logic concepts [6, 8, 7] can only handle acyclic concept descriptions. In addition, the approach for computing the msc proposed by Cohen and Hirsh [8] yields only an approximation of the msc. In fact, since they allow for acyclic descriptions only, they cannot always derive an exact description for the msc. The pragmatic solution proposed in [8] is to restrict the length of value restriction chains occurring in the computed description by some arbitrary but fixed number. This way, one obtains an acyclic description, which may, however, be less specific than the real msc.

Kietz and Morik [9] consider the problem of inductively learning concept descriptions from ABoxes. On the one hand, this work is more restrictive than ours since it does not allow for complex descriptions (not even acyclic ones) in the ABoxes. On the other hand, it tries to solve a more ambitious problem since it tries to learn completely new descriptions from known ABox facts. To this purpose, several heuristic steps are employed. In contrast, computing the lcs and the msc is a purely deductive problem that does not invent new descriptions: it just detects and collects commonalities of given descriptions in an appropriate way.

References

[1] F. Baader. A formal definition for the expressive power of terminological knowledge representation languages. *J. of Logic and Computation*, 6(1):33–54, 1996.

[2] F. Baader. Using automata theory for characterizing the semantics of terminological cycles. *Annals of Mathematics and Artificial Intelligence*, 18(2–4):175–219, 1996.

[3] F. Baader and B. Hollunder. A terminological knowledge representation system with complete inference algorithms. In *Proceedings of the First International Workshop on Processing Declarative Knowledge*, volume 572 of *Lecture Notes in Computer Science*, pages 67–85, Kaiserslautern (Germany), 1991. Springer–Verlag.

[4] F. Baader and U. Sattler. Knowledge representation in process engineering. In *Proceedings of the International Workshop on Description Logics*, Cambridge (Boston), MA, U.S.A., 1996. AAAI Press/The MIT Press.

[5] A. Borgida, R. J. Brachman, D. L. McGuinness, and L. A. Resnick. CLASSIC: A structural data model for objects. In *Proceedings of the 1989 ACM SIGMOD International Conference on Management of Data*, pages 59–67, Portland, OR, 1989.

[6] W. W. Cohen, A. Borgida, and H. Hirsh. Computing least common subsumers in description logics. In William Swartout, editor, *Proceedings of the 10th National Conference on Artificial Intelligence*, pages 754–760, San Jose, CA, July 1992. MIT Press.

[7] W. W. Cohen and H. Hirsh. Learnability of description logics with equality constraints. *Machine Learning*, 17(2/3), 1994.

[8] W. W. Cohen and H. Hirsh. Learning the classic description logic: Theoretical and experimental results. In *Proceedings of KR-94*, pages 121–133, San Francisco, Calif., 1994. Morgan Kaufmann.

[9] J.-U. Kietz and K. Morik. A polynomial approach to the constructive induction of structural knowledge. *Machine Learning Journal*, 14(2):193–218, 1994.

[10] R. Küsters. Characterizing the semantics of terminological cycles in $\mathcal{ALN}$ using finite automata. Technical Report LTCS-97-04, RWTH Aachen, Germany, 1997. ftp://www-lti.informatik.rwth-aachen.de/pub/papers/1997/Kuesters-LTCS-97-04.ps.gz.

[11] R. Küsters. Characterizing the semantics of terminological cycles in $\mathcal{ALN}$ using finite automata. In *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98)*, Trento, Italy, 1998. To appear.

[12] B. Nebel. Terminological cycles: Semantics and computational properties. In J. Sowa, editor, *Formal Aspects of Semantic Networks*, pages 331–361. Morgan Kaufmann, San Mateo, 1991.

[13] W. A. Woods and J. G. Schmolze. The KL-ONE family. *Computers and Mathematics with Applications, special issue on knowledge representation*, 23(2-5):133–177, 1991.