ECAI 2014
T. Schaub et al. (Eds.)

doi:10.3233/978-1-61499-419-0-555

The Logical Difference for $\mathcal{ELH}^r$-Terminologies using Hypergraphs[1]

Michel Ludwig and **Dirk Walther** [2]

Abstract. We propose a novel approach for detecting semantic differences between ontologies. In this paper we investigate the logical difference for $\mathcal{EL}$-terminologies extended with role inclusions, and domain & range restrictions of roles. Three types of queries are covered: concept subsumption, instance and conjunctive queries. Using a hypergraph representation of such ontologies, we show that logical differences can be detected by checking for the existence of simulations between the corresponding hypergraphs. A minor adaptation of the simulation notions allows us to capture different types of queries. We also evaluate our hypergraph approach by applying a prototype implementation on large ontologies.

1 INTRODUCTION

The aim of this paper is to propose and investigate a novel and coherent approach to the logical difference problem as introduced in [6, 8, 9] using a hypergraph representation of ontologies. The logical difference is taken to be the set of queries relevant to an application domain that produce different answers when evaluated over ontologies that are to be compared. The language and signature of the queries can be adapted in such a way that exactly the differences of interest become visible, which can be independent of the syntactic representation of the ontologies. Three types of queries have been studied so far: concept subsumptions, instance and conjunctive queries. The logical difference problem involves reasoning tasks such as determining the existence of a difference and of a succinct representation of the entire set of queries that lead to different answers. Other relevant tasks include the construction of an example query that yields different answers from ontologies given a representation of the difference as well as finding explanations, i.e. the axioms by which this query is entailed.

Our approach is based on representing ontologies as hypergraphs and computing simulations between them. Hypergraphs are a generalisation of graphs with many applications in computer science and discrete mathematics. In knowledge representation hypergraphs have been used implicitly to define reachability-based modules of ontologies [12], explicitly to define locality-based modules [11] and to perform efficient reasoning with ontologies [10]. We consider ontologies that can be translated into directed hypergraphs by taking the signature symbols as nodes and treating the axioms as hyperedges. For instance, the axiom $A \sqsubseteq \exists r.B$ is translated into the hyperedge $(\{x_A\}, \{x_r, x_B\})$, and the axiom $A \equiv B_1 \sqcap B_2$ into the hyperedges $(\{x_A\}, \{x_{B_1}\})$, $(\{x_A\}, \{x_{B_2}\})$ and $(\{x_{B_1}, x_{B_2}\}, \{x_A\})$, where each node x_σ corresponds to the signature symbol σ, respectively. A feature of the translation of axioms into hyperedges is that all the information about the axiom and the logical operators contained in it is preserved. We can treat the terminology and its hypergraph representation interchangeably. The existence of certain simulations for a signature between hypergraphs characterises the fact that the corresponding ontologies cannot be distinguished from each other with queries over the signature, i.e. no logical difference exists. If no simulation can be found, we can directly read off the hypergraph the axioms responsible for the concept inclusion that witnesses the logical difference.

In this paper we follow [6] and consider ontologies formulated as terminologies in the description logic $\mathcal{ELH}^r$, an extension of $\mathcal{EL}$ with role inclusions and domain & range restrictions [1]. Many ontologies are expressed (to a large extent) in the form of such terminologies: the *Systematized Nomenclature of Medicine - Clinical Terms (*SNOMED CT*)* [5] ontology, for example, which now contains definitions for about 300 000 terms, and the *National Cancer Institutes Thesaurus (NCI)* [4] with definitions for about 100 000 terms. Naturally, it is important for ontology engineering to have automated tool support for detecting semantic differences between versions of such large ontologies.

It has been shown that differences between general $\mathcal{ELH}^r$-TBoxes that are observable by instance and conjunctive queries can be detected by concept subsumptions formulated in extensions of $\mathcal{ELH}^r$ [6]. Thus it is sufficient in these cases to consider concept subsumption queries over the extended languages only. *Primitive witness theorems* state that for every concept subsumption in the difference between $\mathcal{ELH}^r$-terminologies, there are also *simpler* subsumptions of the form $A \sqsubseteq D$ or $C \sqsubseteq A$ that have an atomic concept, called *witness*, either on the left-hand or the right-hand side. Checking for the existence of a logical difference is thus equivalent to searching for so-called *left-* and *right-hand witnesses*. In [6] distinct methods based on semantic notions are employed for each type of witness. The search for left-hand witnesses is performed by checking for simulations between canonical models, whereas two different approaches were suggested for right-hand witnesses: one is based on instance checking and the second one employs dynamic programming. In this paper we develop an alternative approach for finding witnesses based on checking for the existence of certain simulations between ontology hypergraphs. The detection of witnesses can be performed by checking for the existence of *forward* and *backward simulations*, which can both be defined independently of semantic notions. Our approach is unifying in the sense that the existence of both types of witnesses can be characterised via graph-theoretic notions. As checking for forward simulations is similar to checking for simulations between canonical models [6], we therefore focus on

[1] This work was supported by the German Research Foundation (DFG) within the Cluster of Excellence 'Center for Advancing Electronics Dresden'.
[2] TU Dresden, Theoretical Computer Science, Dresden, Germany, email: {michel,dirk}@tcs.inf.tu-dresden.de

backward simulations for finding right-hand witnesses in this paper.

We start by reviewing the query languages based on extensions of $\mathcal{EL}$ with role inclusions and domain and range restrictions, $\mathcal{ELH}^r$-terminologies as well as notions related to the logical difference problem. We then introduce the notion of a backward simulation in hypergraphs for potentially *cyclic* $\mathcal{ELH}^r$-terminologies and we show that the existence of backward simulations corresponds to the absence of right-hand witnesses. A prototype implementation of an algorithm that checks for the existence of both types of simulations demonstrates that on acyclic terminologies witnesses can be found almost always quicker than with the previous tool CEX 2.5 [7] in our experiments. Note that the dynamic programming approach for finding right-hand witnesses is implemented in CEX 2.5, which only works for large but acyclic terminologies such as SNOMED CT [6–8]. However, our new implementation can also handle large cyclic terminologies. The instance checking algorithm [6] can handle small cyclic terminologies, but concept subsumptions witnessing a difference cannot be easily constructed using that algorithm.

This paper builds upon results from [6, 8] and extends a previous paper on the concept subsumption difference between terminologies using hypergraphs but restricted to the logic $\mathcal{EL}$ [3].

2 PRELIMINARIES

We start by briefly reviewing the description logic $\mathcal{EL}$ and its extension $\mathcal{EL}^{\mathsf{ran}}$ with range restrictions of roles as well as concept subsumptions based on $\mathcal{ELH}^r$ and $\mathcal{EL}^{\mathsf{ran}}$. For a more detailed introduction to description logics, we refer to [2].

Let $\mathsf{N_C}$ and $\mathsf{N_R}$ be sets of concept names and role names. We assume these sets to be mutually disjoint and countably infinite. Upper-case letters A, B, X, Y, Z denote concept names from $\mathsf{N_C}$, whereas lower-case letters r, s, t represent role names from $\mathsf{N_R}$. The sets of $\mathcal{EL}$*-concepts* C and $\mathcal{EL}^{\mathsf{ran}}$*-concepts* D, and the sets of $\mathcal{ELH}^r$*-inclusions* α and $\mathcal{EL}^{\mathsf{ran}}$*-inclusions* β are built according to the following grammar rules:

$$C ::= \top \mid A \mid C \sqcap C \mid \exists r.C$$
$$D ::= \top \mid A \mid D \sqcap D \mid \exists r.D \mid \mathsf{ran}(r)$$
$$\alpha ::= C \sqsubseteq C \mid \mathsf{ran}(r) \sqsubseteq C \mid \mathsf{ran}(r) \sqcap C \sqsubseteq C \mid r \sqsubseteq s$$
$$\beta ::= D \sqsubseteq C \mid r \sqsubseteq s$$

where $A \in \mathsf{N_C}$ and $r, s \in \mathsf{N_R}$. Concept inclusions of the form $\mathsf{ran}(r) \sqsubseteq D$ or $\mathsf{ran}(r) \sqcap C \sqsubseteq D$ are also called *range restrictions*, and those of the form $\mathsf{dom}(r) \sqsubseteq D$ are termed *domain restrictions*, where $\mathsf{dom}(r)$ stands for $\exists r.\top$. We also refer to such inclusions also as *axioms*. A *TBox* is a finite set of axioms.

The semantics is defined using interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where the domain $\Delta^{\mathcal{I}}$ is a non-empty set, and $\cdot^{\mathcal{I}}$ is a function mapping each concept name A to a subset $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$ and every role name r to a binary relation $r^{\mathcal{I}}$ over $\Delta^{\mathcal{I}}$. The *extension* $C^{\mathcal{I}}$ of a concept C is defined inductively as: $(\top)^{\mathcal{I}} := \Delta^{\mathcal{I}}$, $(C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}}$, $(\exists r.C)^{\mathcal{I}} := \{x \in \Delta^{\mathcal{I}} \mid \exists y \in C^{\mathcal{I}} : (x, y) \in r^{\mathcal{I}}\}$ and $(\mathsf{ran}(r))^{\mathcal{I}} := \{y \in \Delta^{\mathcal{I}} \mid \exists x : (x, y) \in r^{\mathcal{I}}\}$.

An interpretation $\mathcal{I}$ *satisfies* a concept C, an inclusion $C \sqsubseteq D$ or $r \sqsubseteq s$ if, respectively, $C^{\mathcal{I}} \neq \emptyset$, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, or $r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$. Note that every $\mathcal{EL}$- and $\mathcal{EL}^{\mathsf{ran}}$-concept is satisfiable. We write $\mathcal{I} \models \varphi$ if $\mathcal{I}$ satisfies the axiom φ. An interpretation $\mathcal{I}$ *satisfies* a TBox $\mathcal{T}$ if $\mathcal{I}$ satisfies all axioms in $\mathcal{T}$; in this case, we say that $\mathcal{I}$ is a *model* of $\mathcal{T}$. An axiom φ *follows* from a TBox $\mathcal{T}$, written $\mathcal{T} \models \varphi$, if for all models $\mathcal{I}$ of $\mathcal{T}$, we have that $\mathcal{I} \models \varphi$.

An $\mathcal{ELH}^r$*-terminology* $\mathcal{T}$ is an $\mathcal{ELH}^r$-TBox consisting of axioms α of the form $A \sqsubseteq C$, $A \equiv C$, $r \sqsubseteq s$, $\mathsf{ran}(r) \sqsubseteq C$ or $\mathsf{dom}(r) \sqsubseteq C$, where A is a concept name, C an $\mathcal{EL}$-concept and no concept name occurs more than once on the left-hand side of an axiom.[3] To simplify the presentation we assume that terminologies do not contain axioms of the form $A \equiv B$ or $A \equiv \top$ (after having removed multiple $\top$-conjuncts) for concept names A and B. For a terminology $\mathcal{T}$, let $\prec_{\mathcal{T}}$ be a binary relation over $\mathsf{N_C}$ such that $A \prec_{\mathcal{T}} B$ iff there is an axiom of the form $A \sqsubseteq C$ or $A \equiv C$ in $\mathcal{T}$ such that $B \in \mathsf{sig}(C)$. A terminology $\mathcal{T}$ is *acyclic* if the transitive closure $\prec_{\mathcal{T}}^+$ of $\prec_{\mathcal{T}}$ is irreflexive; otherwise $\mathcal{T}$ is *cyclic*. We say that a concept name A is *conjunctive in* $\mathcal{T}$ iff there exist concept names $B_1, \ldots, B_n$, $n > 0$, such that $A \equiv B_1 \sqcap \ldots \sqcap B_n \in \mathcal{T}$; otherwise A is said to be *non-conjunctive in* $\mathcal{T}$. An $\mathcal{ELH}^r$-terminology $\mathcal{T}$ is *normalised* iff it only contains axioms of the forms $r \sqsubseteq s$,

- $\varphi \sqsubseteq B_1 \sqcap \ldots \sqcap B_n$, $A \sqsubseteq \exists r.B$, $A \sqsubseteq \mathsf{dom}(r)$, and
- $A \equiv B_1 \sqcap \ldots \sqcap B_m$, $A \equiv \exists r.B$,

where $\varphi \in \{A, \mathsf{dom}(s), \mathsf{ran}(s)\}$, $n \geq 1$, $m \geq 2$, A, B, B_i are concept names, r, s roles names, and each conjunct B_i is non-conjunctive in $\mathcal{T}$. Every $\mathcal{ELH}^r$-terminology $\mathcal{T}$ can be normalised in polynomial time such that the resulting terminology is a conservative extension of $\mathcal{T}$ [6].

A signature Σ is a finite set of symbols from $\mathsf{N_C}$ and $\mathsf{N_R}$. The signature $\mathsf{sig}(\varphi)$ is the set of concept and role names occurring in φ, where φ ranges over any syntactic object. The symbol Σ is used as a subscript to a set of concepts or inclusions to denote that the elements only use symbols from Σ, e.g., $\mathcal{EL}_\Sigma$, $\mathcal{EL}^{\mathsf{ran}}_\Sigma$, $\mathcal{ELH}^r_\Sigma$, etc.

We now recall some notions of the logical difference from [6,8] for two query languages, $\mathcal{ELH}^r$- and $\mathcal{EL}^{\mathsf{ran}}$-inclusions. The concept inclusion differences for these languages are sufficient to capture concept & instance query differences as defined in [6].

Definition 1 (Concept Inclusion Difference) *Let* $\Gamma \in \{\mathcal{ELH}^r, \mathcal{EL}^{\mathsf{ran}}\}$. *The* Γ-concept inclusion difference *between* $\mathcal{ELH}^r$*-terminologies* $\mathcal{T}_1$ *and* $\mathcal{T}_2$ *w.r.t. a signature* Σ *is the set* $\mathsf{Diff}^\Gamma_\Sigma(\mathcal{T}_1, \mathcal{T}_2)$ *of all* Γ*-inclusions* φ *such that* $\mathsf{sig}(\varphi) \subseteq \Sigma$, $\mathcal{T}_1 \models \varphi$, *and* $\mathcal{T}_2 \not\models \varphi$.

If the set $\mathsf{Diff}^\Gamma_\Sigma(\mathcal{T}_1, \mathcal{T}_2)$ is not empty, then it typically contains infinitely many concept inclusion. We make use of the *primitive witnesses theorems* from [6], which state that if there is a concept inclusion difference in $\mathsf{Diff}^\Gamma_\Sigma(\mathcal{T}_1, \mathcal{T}_2)$, then there exists an inclusion in $\mathsf{Diff}^\Gamma_\Sigma(\mathcal{T}_1, \mathcal{T}_2)$ of one of the following four types $\delta_1, \ldots, \delta_4$, which are built according to the grammar rules below:

$$\begin{aligned} \delta_1 &::= r \sqsubseteq s \\ \delta_2 &::= C \sqsubseteq A \mid \mathsf{ran}(r) \sqcap C \sqsubseteq A \\ \delta_3 &::= D \sqsubseteq A \\ \delta_4 &::= A \sqsubseteq C \mid \mathsf{dom}(r) \sqsubseteq C \mid \mathsf{ran}(r) \sqsubseteq C \end{aligned}$$

where δ_1 ranges over role inclusions, δ_2 and δ_4 are $\mathcal{ELH}^r$-inclusions and δ_3 is an $\mathcal{EL}^{\mathsf{ran}}$-inclusion. Note that each of these inclusions has either a simple left-hand or right-hand side.

The following table summarises results from [6], which identify the types of inclusions that are sufficient to represent the concept inclusion difference between $\mathcal{ELH}^r$-terminologies for the two query languages that we consider (and simply refer to as 'ran' and 'Ran' from now on).[4]

[3] A concept equation $A \equiv C$ stands for the inclusions $A \sqsubseteq C$ and $C \sqsubseteq A$.
[4] We refer to Theorems 40 and 61 in [6].

Query language Γ	ξ	Types of inclusions in $\mathsf{Diff}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)$
$\mathcal{ELH}^r_\Sigma$	ran	δ_1, δ_2, δ_4
$\mathcal{EL}^{\mathsf{ran}}_\Sigma$	Ran	δ_1, δ_3, δ_4

The set of all ξ-concept inclusion difference witnesses is defined as

$$\mathsf{Wtn}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2) = (\mathsf{roleWtn}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2), \mathsf{lhsWtn}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2), \mathsf{rhsWtn}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)),$$

where the set $\mathsf{roleWtn}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)$ consists of all type-δ_1 inclusions in $\mathsf{Diff}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)$, and the sets $\mathsf{lhsWtn}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2) \subseteq (\mathsf{N_C}\cap\Sigma) \cup \{\,\mathsf{dom}(r) \mid r\in\Sigma\,\} \cup \{\,\mathsf{ran}(r) \mid r\in\Sigma\,\}$ and $\mathsf{rhsWtn}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2) \subseteq \mathsf{N_C}\cap\Sigma$ of *left-hand* and *right-hand concept difference witnesses* consist of the left-hand sides of the type-δ_4 inclusions in $\mathsf{Diff}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)$ and the right-hand sides of type-δ_2 and type-δ_3 inclusions in $\mathsf{Diff}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)$, respectively, depending on the query language ξ (cf. table above). Consequently, the set $\mathsf{Wtn}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)$ can be seen as a finite representation of the set $\mathsf{Diff}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)$ [6], which is typically infinite. As a corollary of the primitive witness theorems in [6], we have that the representation is complete in the following sense: $\mathsf{Diff}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)=\emptyset$ iff $\mathsf{Wtn}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)=(\emptyset,\emptyset,\emptyset)$. Thus, to decide the existence of concept inclusion differences, it is equivalent to decide non-emptiness of the three witness sets. In this paper, we focus on right-hand witnesses in $\mathsf{rhsWtn}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)$, i.e., only the inclusions of types δ_2 and δ_3 are relevant.[5]

3 LOGICAL DIFFERENCE USING HYPERGRAPHS

Our approach for detecting logical differences is based on finding appropriate simulations between the hypergraph representations of terminologies. The hypergraph notion in this paper is such that the existence of certain simulations between the ontology hypergraphs of terminologies $\mathcal{T}_1$ and $\mathcal{T}_2$ coincides with $\mathsf{Wtn}^{\xi}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)=(\emptyset,\emptyset,\emptyset)$, irrespective of the query language $\xi\in\{\mathsf{ran},\mathsf{Ran}\}$ used to detect differences. The simulation notions are defined in such a way that the different query languages are taken into account. For every concept name A in Σ (or role name $r\in\Sigma$), one has to check whether A (or $\mathsf{dom}(r)$, $\mathsf{ran}(r)$) belongs to $\mathsf{lhsWtn}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)$ or to $\mathsf{rhsWtn}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)$. For the former, we check for the existence of a *forward simulation*, and for the latter, for the existence of a *backward simulation* between the ontology hypergraphs of $\mathcal{T}_1$ and $\mathcal{T}_2$. In this paper we present backward simulations only. Checking for left-hand side witnesses as well as for witnesses in $\mathsf{roleWtn}_{\Sigma}(\mathcal{T}_1,\mathcal{T}_2)$ using ontology hypergraphs can be done similarly to [6].

We now define ontology hypergraphs and we introduce the notion of a backward simulation between such hypergraphs. Subsequently, we analyse the computational complexity of checking for the existence of backward simulations.

3.1 Ontology Hypergraphs

We introduce a hypergraph representation of $\mathcal{ELH}^r$-terminologies.

A *directed hypergraph* is a tuple $\mathcal{G}=(\mathcal{V},\mathcal{E})$, where $\mathcal{V}$ is a non-empty set of *nodes* (or *vertices*), and $\mathcal{E}$ is a set of *directed hyperedges* of the form $e=(S,S')$, where $S,S'\subseteq\mathcal{V}$. We now show how to represent terminologies as hypergraphs.

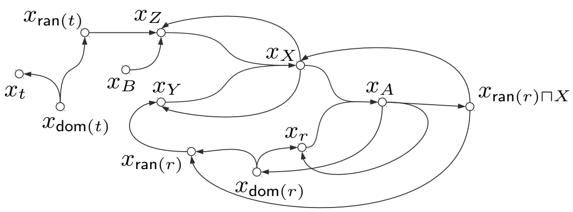

Figure 1. Example Hypergraph

Definition 2 (Ontology Hypergraph) *Let $\mathcal{T}$ be a normalised $\mathcal{ELH}^r$-terminology and let Σ be a signature. The* ontology hypergraph $\mathcal{G}^{\Sigma}_{\mathcal{T}}$ of $\mathcal{T}$ for Σ *is a directed hypergraph $\mathcal{G}^{\Sigma}_{\mathcal{T}}=(\mathcal{V},\mathcal{E})$ defined as follows:*

$$\begin{aligned}
\mathcal{V} = {} & \{\, x_A \mid A\in\mathsf{N_C}\cap(\Sigma\cup\mathsf{sig}(\mathcal{T}))\,\} \\
& \cup \{\, x_r, x_{\mathsf{dom}(r)}, x_{\mathsf{ran}(r)} \mid r\in\mathsf{N_R}\cap(\Sigma\cup\mathsf{sig}(\mathcal{T}))\,\} \\
& \cup \{\, x_{\mathsf{ran}(r)\sqcap B} \mid A\bowtie\exists r.B\in\mathcal{T},\ \bowtie\in\{\sqsubseteq,\equiv\}\,\}, \textit{and} \\
\mathcal{E} = {} & \{(\{x_\ell\},\{x_{B_i}\}) \mid \ell\bowtie B_1\sqcap\ldots\sqcap B_n\in\mathcal{T},\ \bowtie\in\{\sqsubseteq,\equiv\}, \\
& \qquad \ell\in\mathsf{N_C}\cup\{\,\mathsf{dom}(s),\mathsf{ran}(s)\mid s\in\mathsf{N_R}\,\}\} \\
& \cup \{(\{x_A\},\{x_{\mathsf{dom}(r)}\}) \mid A\sqsubseteq\mathsf{dom}(r)\in\mathcal{T} \textit{ or } A\bowtie\exists r.B\in\mathcal{T}, \\
& \qquad \bowtie\in\{\sqsubseteq,\equiv\}\} \\
& \cup \{(\{x_A\},\{x_r,x_{\mathsf{ran}(r)\sqcap B}\}), (\{x_{\mathsf{ran}(r)\sqcap B}\},\{x_B\}), \\
& \qquad (\{x_{\mathsf{ran}(r)\sqcap B}\},\{x_{\mathsf{ran}(r)}\}) \mid A\bowtie\exists r.B\in\mathcal{T},\ \bowtie\in\{\sqsubseteq,\equiv\}\} \\
& \cup \{\,(\{x_r,x_B\},\{x_A\}) \mid A\equiv\exists r.B\in\mathcal{T}\,\} \\
& \cup \{\,(\{x_{B_1},\ldots,x_{B_n}\},\{x_A\}) \mid A\equiv B_1\sqcap\ldots\sqcap B_n\in\mathcal{T}\,\} \\
& \cup \{\,(\{x_{\mathsf{dom}(r)}\},\{x_r,x_{\mathsf{ran}(r)}\}) \mid r\in\mathsf{N_R}\cap(\Sigma\cup\mathsf{sig}(\mathcal{T})\,\} \\
& \cup \{(\{x_r\},\{x_s\}), (\{x_{\mathsf{dom}(r)}\},\{x_{\mathsf{dom}(s)}\}), \\
& \qquad (\{x_{\mathsf{ran}(r)}\},\{x_{\mathsf{ran}(s)}\}) \mid r\sqsubseteq s\in\mathcal{T}\}
\end{aligned}$$

The ontology hypergraph $\mathcal{G}^{\Sigma}_{\mathcal{T}}=(\mathcal{V},\mathcal{E})$ contains a node x_ℓ for every signature symbol ℓ in Σ and $\mathcal{T}$.[6] Additionally, we represent concepts of the form $\mathsf{dom}(r)$, $\mathsf{ran}(r)$ and $\mathsf{ran}(r)\sqcap B$ as nodes in the graph. We include a hyperedge $(\{x_{\mathsf{dom}(r)}\},\{x_r,x_{\mathsf{ran}(r)}\})$ for every role name r in Σ and $\mathcal{T}$, which corresponds to the tautology $\mathsf{dom}(r)\sqsubseteq\exists r.\mathsf{ran}(r)$.[7] Recall that $\mathsf{dom}(r)$ equals $\exists r.\top$. The other hyperedges in $\mathcal{G}^{\Sigma}_{\mathcal{T}}$ represent the axioms in $\mathcal{T}$. Every hyperedge is directed and can be understood as an implication, i.e., $(\{x_{\ell_1}\},\{x_{\ell_2}\})$ stands for $\mathcal{T}\models\ell_1\sqsubseteq\ell_2$. The complex hyperedges are of the form $(\{x_A\},\{x_r,x_B\})$ and $(\{x_r,x_B\},\{x_A\})$ representing $\mathcal{T}\models A\sqsubseteq\exists r.B$ and $\mathcal{T}\models\exists r.B\sqsubseteq A$, and of the form $(\{x_{B_1},...,x_{B_n}\},\{x_A\})$ representing $\mathcal{T}\models B_1\sqcap\ldots\sqcap B_n\sqsubseteq A$.

Example 3 *Let $\mathcal{T}=\{A\equiv\exists r.X,\ X\equiv Y\sqcap Z,\ B\sqsubseteq Z,\ \mathsf{ran}(r)\sqsubseteq Y,\ \mathsf{ran}(t)\sqsubseteq Z\}$, and $\Sigma=\{A,B,r,t\}$. Then the ontology hypergraph $\mathcal{G}^{\Sigma}_{\mathcal{T}}$ of $\mathcal{T}$ for Σ is depicted in Figure 1. Note that hyperedges of the form $(\{x_Y,x_Z\},\{x_X\})$ are depicted using merging arrows from x_Y and x_Z to x_X.*

In the following we write $\{x_l\}\to_{\mathcal{T}}\{x_{l'}\}$, for $x_l,x_{l'}\in\mathcal{V}$, to denote that $\mathcal{T}\models l\sqsubseteq l'$ holds. Note that such a relation $\to_{\mathcal{T}}$ can be defined directly on ontology hypergraphs as a special reachability relation [3].

3.2 Backward Simulations

We now introduce backward simulations between ontology hypergraphs whose existence coincides with the absence of right-hand

[5] Note that type-δ_3 inclusions are sufficient to capture the right-hand witnesses w.r.t. an extension of $\mathcal{EL}^{\mathsf{ran}}$-inclusions that serves as query language suitable for conjunctive queries [6].

[6] Note that, differently to [3], the graph $\mathcal{G}^{\Sigma}_{\mathcal{T}}$ does not contain any node representing $\top$.

[7] These hyperedges are relevant for forward simulations only.

witnesses. The simulations are defined in such a way that a node x_A in $\mathcal{G}^\Sigma_{\mathcal{T}_1}$ is simulated by a node $x_{A'}$ in $\mathcal{G}^\Sigma_{\mathcal{T}_2}$ iff A' is entailed in $\mathcal{T}_2$ by exactly the same Σ-concepts that entail A in $\mathcal{T}_1$, depending on the query language (as it is made precise in Thm. 1 below). To define backward simulations we need to take all the axioms that cause Σ-concepts to entail a concept name into account. Axioms of the forms $A \equiv \exists r.B$, $A \equiv B_1 \sqcap \ldots \sqcap B_n$, and $\mathsf{ran}(r) \sqsubseteq A$ require special treatment, while it is more straightforward to deal with the other types of axioms. For the former, consider $\mathcal{T}_1 = \{A \equiv \exists r.X\}$, $\mathcal{T}_2 = \{A \sqsubseteq \exists r.\top\}$, and $\Sigma = \{A, r\}$. It holds that $\mathsf{Diff}^\xi_\Sigma(\mathcal{T}_1, \mathcal{T}_2) = \emptyset$ with $\xi \in \{\mathsf{ran}, \mathsf{Ran}\}$. Observe that there does not exist a Σ-concept in any of the query languages that entails A in $\mathcal{T}_1$ as the concept name X is not entailed by any Σ-concept in $\mathcal{T}_1$. Thus, the node x_A in $\mathcal{G}^\Sigma_{\mathcal{T}_1}$ should be simulated by the node x_A in $\mathcal{G}^\Sigma_{\mathcal{T}_2}$. To handle such cases, we want to characterise the entailment by a Σ-concept formulated in any of the query languages in terms of (a special) reachability notion, named Σ-*reachability*, in ontology hypergraphs.

We need to take special care of axioms of the form $\mathsf{ran}(r) \sqsubseteq X$ as they might cause non-obvious entailments. Let $\mathcal{T} = \{X \equiv B_1 \sqcap B_2,\ A \equiv \exists r.X\}$ and $\Sigma = \{A, B_1, B_2, r\}$. Then the Σ-concept $\exists r.(B_1 \sqcap B_2)$ entails A in $\mathcal{T}$. If we add the axiom $\mathsf{ran}(r) \sqsubseteq B_1$ to $\mathcal{T}$, then already the Σ-concept $\exists r.B_2$ (of smaller signature) is sufficient to entail A in $\mathcal{T}$. Intuitively, the conjunct B_1 of X is already covered by $\mathsf{ran}(r)$ in the presence of the axiom $\mathsf{ran}(r) \sqsubseteq B_1$ (as $\mathcal{T} \models \mathsf{ran}(r) \sqcap B_2 \sqsubseteq X$). To define backward simulations for axioms of the form $A \equiv \exists r.X$, all axioms of the form $\mathsf{ran}(r) \sqsubseteq Y$ need to be taken into account. We therefore define the notion of Σ-reachability using an additional parameter $\zeta \in \{\epsilon\} \cup (\mathsf{N_R} \cap \Sigma)$ which we call the *context of a role*, i.e. an expression of the form $\mathsf{ran}(\zeta)$. We treat ϵ as a special role name and set $\mathsf{ran}(\epsilon) = \top$. The set of all role contexts, in symbols $\mathcal{C}^\Sigma$, is defined as $\mathcal{C}^\Sigma = \{\epsilon\} \cup (\mathsf{N_R} \cap \Sigma)$.

For a signature Σ, let $\Sigma^{\mathsf{dom}} = \{\, \mathsf{dom}(t) \mid t \in \mathsf{N_R} \cap \Sigma \,\}$ and $\Sigma^{\mathsf{ran}} = \{\, \mathsf{ran}(t) \mid t \in \mathsf{N_R} \cap \Sigma \,\}$ be the sets consisting of concepts of the form $\mathsf{dom}(t)$ and $\mathsf{ran}(t)$ for every role name t in Σ, respectively. Furthermore let $\Sigma^{(\mathsf{ran},\zeta)} = \Sigma \cup \Sigma^{\mathsf{dom}} \cup \{\, \mathsf{ran}(\zeta) \mid \zeta \neq \epsilon \,\}$ and $\Sigma^{(\mathsf{Ran},\zeta)} = \Sigma \cup \Sigma^{\mathsf{dom}} \cup \Sigma^{\mathsf{ran}}$, for $\zeta \in \mathcal{C}^\Sigma$. Note that $\Sigma^{(\mathsf{Ran},\zeta)} = \Sigma^{(\mathsf{Ran},\zeta')}$ for every $\zeta, \zeta' \in \mathcal{C}^\Sigma$.

Definition 4 (Σ-Reachability) *Let $\mathcal{G}^\Sigma_\mathcal{T} = (\mathcal{V}, \mathcal{E})$ be the ontology hypergraph of a normalised $\mathcal{ELH}^r$-terminology $\mathcal{T}$ for a signature Σ. The sets $\mathcal{V}_{(\xi,\zeta)} \subseteq \mathcal{V}$ for $\xi \in \{\mathsf{ran}, \mathsf{Ran}\}$ and $\zeta \in \mathcal{C}^\Sigma$ are inductively defined as:*

- (i^ξ_e) *$x \in \mathcal{V}_{(\xi,\zeta)}$ if $\{x_\sigma\} \to_\mathcal{T} \{x\}$ for some $\sigma \in \Sigma^{(\xi,\zeta)}$;*
- (ii^{ran}_e) *$x \in \mathcal{V}_{(\mathsf{ran},\zeta)}$ if $(\{x_s, y\}, \{x\}) \in \mathcal{E}$ with $\{x_s, y\} \subseteq \mathcal{V}_{(\mathsf{ran},s)}$;*
- (ii^{Ran}_e) *$x \in \mathcal{V}_{(\mathsf{Ran},\zeta)}$ if $(\{x_r, y\}, \{x\}) \in \mathcal{E}$ with $\{x_r, y\} \subseteq \mathcal{V}_{(\mathsf{Ran},\zeta)}$;*
- (iii^ξ_e) *$x \in \mathcal{V}_{(\xi,\zeta)}$ if $(\{x_{X_1}, \ldots, x_{X_n}\}, \{x\}) \in \mathcal{E}$ and $\{x_{X_1}, ..., x_{X_n}\} \subseteq \mathcal{V}_{(\xi,\zeta)}$.*

A node $x \in \mathcal{V}$ is (Σ, ξ, ζ)-reachable *in $\mathcal{G}^\Sigma_\mathcal{T}$ iff $x \in \mathcal{V}_{(\xi,\zeta)}$.*

Note that different conditions apply for $\xi = \mathsf{ran}$ and $\xi = \mathsf{Ran}$. The role context ζ is irrelevant for $(\Sigma, \mathsf{Ran}, \zeta)$-reachability, whereas it is essential for $(\Sigma, \mathsf{ran}, \zeta)$-reachability as illustrated in the following example.

Example 5 *Let $\mathcal{T} = \{A \equiv \exists r.X,\ X \equiv Y \sqcap Z,\ B \sqsubseteq Z,\ \mathsf{ran}(r) \sqsubseteq Y,\ \mathsf{ran}(t) \sqsubseteq Z\}$ (cf. Example 3) and let $\Sigma = \{B, r\}$. All nodes but x_t, $x_{\mathsf{dom}(t)}$, $x_{\mathsf{ran}(t)}$ and $x_{\mathsf{ran}(r) \sqcap X}$ are (Σ, ξ, r)-reachable in $\mathcal{G}^\Sigma_\mathcal{T}$ for $\xi \in \{\mathsf{ran}, \mathsf{Ran}\}$, and $(\Sigma, \mathsf{Ran}, \epsilon)$-reachable in $\mathcal{G}^\Sigma_\mathcal{T}$, but only the nodes x_A, x_B, x_Z, x_r and $x_{\mathsf{dom}(r)}$ are $(\Sigma, \mathsf{ran}, \epsilon)$-reachable in $\mathcal{G}^\Sigma_\mathcal{T}$. Note that x_A is $(\Sigma, \mathsf{ran}, \epsilon)$-reachable due to Condition (ii^{ran}_e) and the fact that x_X and x_r are $(\Sigma, \mathsf{ran}, r)$-reachable. We have that $\mathcal{T} \models \exists r.B \sqsubseteq A$.*

To compute all the nodes in a given graph $\mathcal{G}_\mathcal{T}$ that are (Σ, ξ, ζ)-reachable, one can proceed as follows. In a first step one identifies all the nodes x that fulfill Condition (i^ξ_e) by using the relation $\to_\mathcal{T}$. Subsequently, the status of being (Σ, ξ, ζ)-reachable can be propagated appropriately (depending on the context) to the other nodes using the other conditions. It can be readily seen that these computation steps can be performed in polynomial time.

We now state the desired properties of the notion of Σ-reachability.

Lemma 6 *Let $\mathcal{T}$ be a normalised $\mathcal{ELH}^r$-terminology and Σ be a signature. For $\xi \in \{\mathsf{ran}, \mathsf{Ran}\}$, $\zeta \in \mathsf{N_R} \cup \{\epsilon\}$, $A \in \mathsf{N_C}$ and $s \in \mathsf{N_R}$, the following statements hold:*

- (i) *$x_A \in \mathcal{V}$ is $(\Sigma, \mathsf{ran}, \zeta)$-reachable in $\mathcal{G}^\Sigma_\mathcal{T}$ iff there is an $\mathcal{EL}_\Sigma$-concept C such that $\mathcal{T} \models \mathsf{ran}(\zeta) \sqcap C \sqsubseteq A$;*
- (ii) *$x_A \in \mathcal{V}$ is $(\Sigma, \mathsf{Ran}, \zeta)$-reachable in $\mathcal{G}^\Sigma_\mathcal{T}$ iff there is an $\mathcal{EL}^{\mathsf{ran}}_\Sigma$-concept D such that $\mathcal{T} \models D \sqsubseteq A$;*
- (iii) *$x_s \in \mathcal{V}$ is (Σ, ξ, ζ)-reachable in $\mathcal{G}^\Sigma_\mathcal{T}$ iff there exists $s' \in \mathsf{N_R} \cap \Sigma$ such that $\mathcal{T} \models s' \sqsubseteq s$.*

For axioms of the form $A \equiv B_1 \sqcap \ldots \sqcap B_n$, we introduce the following notion which associates with every node x_A in a hypergraph $\mathcal{G}_\mathcal{T}$ a set of concept names non-conj(x_A) that are essential to entail A in $\mathcal{T}$ (also see [6] for a similar notion).

Definition 7 (Non-Conjunctive) *Let $\mathcal{G}^\Sigma_\mathcal{T} = (\mathcal{V}, \mathcal{E})$ be the ontology hypergraph of a normalised $\mathcal{ELH}^r$-terminology $\mathcal{T}$ for a signature Σ. For $x_A \in \mathcal{V}$, let non-conj(x_A) be defined as follows:*

- *if $(\{x_{B_1}, \ldots, x_{B_n}\}, \{x_A\}) \in \mathcal{E}$ (i.e. $n \geq 2$ and $A \equiv B_1 \sqcap \ldots \sqcap B_n \in \mathcal{T}$), we set non-conj$_\mathcal{T}(x_A) = \{x_{B_1}, \ldots, x_{B_n}\}$;*
- *otherwise, let non-conj$_\mathcal{T}(x_A) = \{x_A\}$.*

We say that a node $y \in \mathcal{V}$ is relevant *for a node x in $\mathcal{T}$* w.r.t. *a set of* node labels *$\mathcal{L}$ used in $\mathcal{G}^\Sigma_\mathcal{T}$ if $y \in$ non-conj$_\mathcal{T}(x)$ and $\{x_\ell\} \not\to_\mathcal{T} \{y\}$ for every $\ell \in \mathcal{L}$.*

We note that, for $\mathcal{G}^\Sigma_\mathcal{T} = (\mathcal{V}, \mathcal{E})$, it holds that $(\{x_{B_1}, \ldots, x_{B_n}\}, \{x_A\}) \in \mathcal{E}$ iff $A \equiv B_1 \sqcap \ldots \sqcap B_n \in \mathcal{T}$. Hence, for every $\zeta \in \{\epsilon\} \cup (\mathsf{N_R} \cap \Sigma)$ and for every $\mathcal{EL}_\Sigma$-concept C it holds that $\mathcal{T} \models \mathsf{ran}(\zeta) \sqcap C \sqsubseteq A$ iff for every $x_X \in$ non-conj$_\mathcal{T}(x_A)$, either $\mathcal{T} \models \mathsf{ran}(\zeta) \sqsubseteq X$ (if $\zeta \neq \epsilon$) or $\mathcal{T} \models C \sqsubseteq X$.

We now define *backward simulations* as subsets of $\mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{C}^\Sigma$. The elements of $\mathcal{C}^\Sigma$ represent the role context (an expression of the form $\mathsf{ran}(\zeta)$ with $\zeta \in \mathcal{C}^\Sigma$) in which a node $x \in \mathcal{V}_1$ should be simulated by a node $x' \in \mathcal{V}_2$.

Definition 8 (Backward Simulation) *Let $\mathcal{G}^\Sigma_{\mathcal{T}_1} = (\mathcal{V}_1, \mathcal{E}_1)$, $\mathcal{G}^\Sigma_{\mathcal{T}_2} = (\mathcal{V}_2, \mathcal{E}_2)$ be the ontology hypergraphs of normalised $\mathcal{ELH}^r$-terminologies $\mathcal{T}_1$ and $\mathcal{T}_2$ for a signature Σ. For $\xi \in \{\mathsf{ran}, \mathsf{Ran}\}$, a relation $\hookleftarrow^\xi \subseteq \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{C}^\Sigma$ is a* backward (Σ, ξ)-simulation *if the following conditions are satisfied:*

- (i^ξ_b) *if $(x, x', \zeta) \in \hookleftarrow^\xi$, then for every $\sigma \in \Sigma^{(\xi,\zeta)}$: $\{x_\sigma\} \to_{\mathcal{T}_1} \{x\}$ implies $\{x_\sigma\} \to_{\mathcal{T}_2} \{y'\}$ for every $y' \in \mathcal{V}_2$ relevant for x' in $\mathcal{T}_2$ w.r.t. $\{\mathsf{ran}(\zeta)\}$;*
- (ii^ξ_b) *if $(x, x', \zeta) \in \hookleftarrow^\xi$ and $(\{x_r, y\}, \{x\}) \in \mathcal{E}_1$, then for every $s \in \Sigma$ such that $\{x_s\} \to_{\mathcal{T}_1} \{x_r\}$ and y is (Σ, ξ, s)-reachable, and for every $y' \in \mathcal{V}_2$ relevant for x' in $\mathcal{T}_2$ w.r.t. $\{\mathsf{ran}(\zeta), \mathsf{dom}(s)\}$ there exists $(\{x_{r'}, z'\}, \{y'\}) \in \mathcal{E}_2$ with $\{x_s\} \to_{\mathcal{T}_2} \{x_{r'}\}$ and $(y, z', s) \in \hookleftarrow^\xi$;*
- (iii^ξ_b) *if $(x, x', \zeta) \in \hookleftarrow^\xi$ and $(\{x_{X_1}, \ldots, x_{X_n}\}, \{x\}) \in \mathcal{E}_1$, then for every $y' \in \mathcal{V}_2$ relevant for x' w.r.t. $\{\mathsf{ran}(\zeta)\}$ there exists $y \in \mathcal{V}_1$ relevant for x in $\mathcal{T}_1$ w.r.t. $\{\mathsf{ran}(\zeta)\}$ with $(y, y', \epsilon) \in \hookleftarrow^\xi$.*

We write $\mathcal{G}^{\Sigma}_{\mathcal{T}_1} \hookleftarrow^{\mathsf{ran}} \mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ iff there exists a backward (Σ, ran)-simulation $\hookleftarrow^{\mathsf{ran}} \subseteq \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{C}^{\Sigma}$ such that $(x_A, x_A, \epsilon) \in \hookleftarrow^{\mathsf{ran}}$ and $(x_A, x_A, r) \in \hookleftarrow^{\mathsf{ran}}$ for every $A, r \in \Sigma$, and $\mathcal{G}^{\Sigma}_{\mathcal{T}_1} \hookleftarrow^{\mathsf{Ran}} \mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ iff there exists a backward (Σ, Ran)-simulation $\hookleftarrow^{\mathsf{Ran}} \subseteq \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{C}^{\Sigma}$ such that $(x_A, x_A, \epsilon) \in \hookleftarrow^{\mathsf{Ran}}$ for every $A \in \mathsf{N_C} \cap \Sigma$.

Members of a backward simulation $\hookleftarrow^{\xi}$ are called *simulation triples*.

For a node x in $\mathcal{G}^{\Sigma}_{\mathcal{T}_1}$ to be backward simulated by x' in $\mathcal{G}^{\Sigma}_{\mathcal{T}_2}$, Conditions (i_b^{ran}) and (i_b^{Ran}) enforce that appropriate Σ-concept names B or concepts of the form $\mathsf{ran}(s)$, $\mathsf{dom}(s)$ with $s \in \Sigma$ that entail x in $\mathcal{T}_1$ must also entail x' in $\mathcal{T}_2$. Conditions (ii_b^{ran}) and (ii_b^{Ran}) apply to nodes $x_A \in \mathcal{G}^{\Sigma}_{\mathcal{T}_1}$ for which there exists an axiom $A \equiv \exists r.X$ in $\mathcal{T}_1$ and propagate the simulation to the successor node x_X by taking into account possible entailments regarding domain or range restrictions in $\mathcal{T}_2$. Condition (iii_b^{ran}) and (iii_b^{Ran}) handle axioms of the form $A \equiv B_1 \sqcap \ldots \sqcap B_n$ in $\mathcal{T}_1$. We have to match every conjunct y' that is relevant for x' in $\mathcal{T}_2$ with some conjunct y relevant for x in $\mathcal{T}_1$ (possibly leaving some conjuncts y unmatched) since, intuitively speaking, some conjuncts in the definition of A in $\mathcal{T}_1$ can be ignored to preserve logical entailment. For instance, let $\mathcal{T}_1 = \{A \equiv B_1 \sqcap B_2\}$, $\mathcal{T}_2 = \{B_1 \sqsubseteq A\}$ and $\Sigma = \{A, B_1, B_2\}$. Then $\mathsf{rhsWtn}^{\mathsf{ran}}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2) = \emptyset$ and, in particular, $\mathcal{T}_2 \models B_1 \sqcap B_2 \sqsubseteq A$ holds as well. Note that the simulation between conjuncts is propagated in the context ϵ only as all the conjuncts that are entailed by $\mathsf{ran}(\zeta)$ have been filtered out already.

Example 9 *Let $\mathcal{T}_1 = \{A \equiv \exists r.X,\ X \equiv Y \sqcap Z,\ B \sqsubseteq Z,\ \mathsf{ran}(r) \sqsubseteq Y,\ \mathsf{ran}(t) \sqsubseteq Z\}$ (cf. Example 3), $\mathcal{T}_2 = \{A \equiv X \sqcap Y,\ X \equiv \exists r.B,\ \mathsf{dom}(s) \sqsubseteq Y,\ r \sqsubseteq s\}$, and $\Sigma = \{A, B, r, t\}$.*

It can be readily seen that the nodes x_B, x_Y, x_Z, and x_X are $(\Sigma, \mathsf{ran}, r)$-reachable in $\mathcal{G}^{\Sigma}_{\mathcal{T}_1}$. As only $\{x_B\} \to_{\mathcal{T}_1} \{x_B\}$, we have that the node x_B in $\mathcal{G}^{\Sigma}_{\mathcal{T}_1}$ can be simulated by the node x_B in $\mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ in the contexts ϵ, r, and t. Similarly, as only $\{x_B\} \to_{\mathcal{T}_1} \{x_Z\}$, the node x_Z in $\mathcal{G}^{\Sigma}_{\mathcal{T}_1}$ can be simulated by the node x_B in $\mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ in the context ϵ, r but not in t. Hence, as non-conj$_{\mathcal{T}_2}(x_B) = \{x_B\}$ and as x_Z is relevant for x_X in $\mathcal{T}_1$ w.r.t. $\{\mathsf{ran}(r)\}$, we have that x_X in $\mathcal{G}^{\Sigma}_{\mathcal{T}_1}$ can be simulated by x_B in $\mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ in the context r. Finally, as non-conj$_{\mathcal{T}_2}(x_A) = \{x_X, x_Y\}$ and as only x_X is relevant for x_A in $\mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ (due to $\{x_{\mathsf{dom}(r)}\} \to_{\mathcal{T}_2} \{x_Y\}$), we conclude that the node x_A in $\mathcal{G}^{\Sigma}_{\mathcal{T}_1}$ can be simulated by x_A in $\mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ in the contexts ϵ, r, t. Overall,

$$\begin{aligned} S = &\{\, (x_A, x_A, \zeta) \mid \zeta \in \{\epsilon, r, t\} \,\} \cup \{\, (x_B, x_B, \zeta) \mid \zeta \in \{\epsilon, r, t\} \,\} \\ &\cup \{(x_Z, x_B, \epsilon), (x_Z, x_B, r), (x_X, x_B, r)\} \end{aligned}$$

is a backward (Σ, ran)-simulation between $\mathcal{G}^{\Sigma}_{\mathcal{T}_1}$ and $\mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ such that $(x_{A'}, x_{A'}, \epsilon) \in S$ and $(x_{A'}, x_{A'}, r') \in S$ for every $A', r' \in \Sigma$.

Example 10 *Let $\mathcal{T}_1$, $\mathcal{T}_2$, and Σ be defined as in Example 9. We observe that $\{x_{\mathsf{ran}(t)}\} \to_{\mathcal{T}_2} \{x'\}$ does not hold for any node $x' \in \mathcal{G}^{\Sigma}_{\mathcal{T}_2}$, i.e., the node x_Z in $\mathcal{G}^{\Sigma}_{\mathcal{T}_1}$ cannot be Ran-simulated by any node in $\mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ (in any context) as Condition (i_b^{Ran}) cannot be fulfilled. Hence, the node x_X in $\mathcal{G}^{\Sigma}_{\mathcal{T}_1}$ cannot be simulated by x_B in $\mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ in the context r as Condition (iii_b^{Ran}) is violated. Thus, there cannot exist a backward (Σ, Ran)-simulation such that x_A in $\mathcal{G}^{\Sigma}_{\mathcal{T}_1}$ is simulated by x_A in $\mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ in the context ϵ as Condition (ii_b^{Ran}) cannot be fulfilled.*

We now show that the existence of a backward simulation coincides with the absence of right-hand witnesses.

Lemma 11 *Let $\mathcal{T}_1, \mathcal{T}_2$ be normalised $\mathcal{ELH}^r$-terminologies, and let Σ be a signature. For $\xi \in \{\mathsf{ran}, \mathsf{Ran}\}$, if $\mathcal{G}^{\Sigma}_{\mathcal{T}_1} \hookleftarrow^{\xi} \mathcal{G}^{\Sigma}_{\mathcal{T}_2}$, then it holds that:*

(i) $\xi = \mathsf{ran}$: for every $\mathcal{EL}_{\Sigma}$-concept C and for every triple $(x_A, x_{A'}, \zeta) \in \hookleftarrow^{\mathsf{ran}}$, $\mathcal{T}_1 \models \mathsf{ran}(\zeta) \sqcap C \sqsubseteq A$ implies $\mathcal{T}_2 \models \mathsf{ran}(\zeta) \sqcap C \sqsubseteq A'$;

(ii) $\xi = \mathsf{Ran}$: for every $\mathcal{EL}^{\mathit{ran}}_{\Sigma}$-concept D and for every triple $(x_A, x_{A'}, \zeta) \in \hookleftarrow^{\mathsf{Ran}}$, $\mathcal{T}_1 \models \mathsf{ran}(\zeta) \sqcap D \sqsubseteq A$ implies $\mathcal{T}_2 \models \mathsf{ran}(\zeta) \sqcap D \sqsubseteq A'$.

Lemma 12 *Let $\mathcal{T}_1, \mathcal{T}_2$ be normalised $\mathcal{ELH}^r$-terminologies, and let Σ be a signature. Then for $\xi \in \{\mathsf{ran}, \mathsf{Ran}\}$: $\mathsf{rhsWtn}^{\xi}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2) = \emptyset$ implies $\mathcal{G}^{\Sigma}_{\mathcal{T}_1} \hookleftarrow^{\xi} \mathcal{G}^{\Sigma}_{\mathcal{T}_2}$.*

We now obtain the following theorem.

Theorem 1 *Let $\mathcal{T}_1, \mathcal{T}_2$ be normalised $\mathcal{ELH}^r$-terminologies, and let Σ be a signature. Then for $\xi \in \{\mathsf{ran}, \mathsf{Ran}\}$: $\mathsf{rhsWtn}^{\xi}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2) = \emptyset$ iff $\mathcal{G}^{\Sigma}_{\mathcal{T}_1} \hookleftarrow^{\xi} \mathcal{G}^{\Sigma}_{\mathcal{T}_2}$.*

3.3 Computational Complexity

Given two ontology hypergraphs $\mathcal{G}^{\Sigma}_{\mathcal{T}_1} = (\mathcal{V}_1, \mathcal{E}_1)$ and $\mathcal{G}^{\Sigma}_{\mathcal{T}_2} = (\mathcal{V}_2, \mathcal{E}_2)$, one can use the following elimination procedure to check whether $\mathcal{G}^{\Sigma}_{\mathcal{T}_1} \hookleftarrow^{\xi} \mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ holds for $\xi \in \{\mathsf{ran}, \mathsf{Ran}\}$. First, let $S_0^{\xi} \subseteq \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{C}^{\Sigma}$ be the set of all the triples that fulfill Conditions (i_b^{ξ}). Subsequently, we iterate over the elements contained in the set S_i^{ξ} and remove those pairs which do not satisfy Conditions (ii_b^{ξ}) or (iii_b^{ξ}) to obtain the set S_{i+1}^{ξ}. This process terminates with $S_j^{\xi} = S_{j+1}^{\xi}$ for some index j. It holds that $\mathcal{G}^{\Sigma}_{\mathcal{T}_1} \hookleftarrow^{\mathsf{Ran}} \mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ iff $(x_A, x_A, \epsilon) \in S_j^{\mathsf{Ran}}$ for every $A \in \mathsf{N_C} \cap \Sigma$, or that $\mathcal{G}^{\Sigma}_{\mathcal{T}_1} \hookleftarrow^{\mathsf{ran}} \mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ iff $(x_A, x_A, \epsilon) \in S_j^{\mathsf{ran}}$ and $(x_A, x_A, r) \in S_j^{\mathsf{ran}}$ for every $A, r \in \Sigma$.

The procedure described above terminates in at most $|\mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{C}^{\Sigma}|$ iterations. We hence obtain the following theorem.

Theorem 2 *Let $\mathcal{T}_1$, $\mathcal{T}_2$ be normalised $\mathcal{ELH}^r$-terminologies and let Σ be a signature. Then it can be checked in polynomial time whether $\mathcal{G}^{\Sigma}_{\mathcal{T}_1} \hookleftarrow^{\xi} \mathcal{G}^{\Sigma}_{\mathcal{T}_2}$ holds for $\xi \in \{\mathsf{ran}, \mathsf{Ran}\}$.*

3.4 Computing Difference Examples

So far we have focused on finding concept names A belonging to the set $\mathsf{rhsWtn}^{\xi}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2)$ with $\xi \in \{\mathsf{ran}, \mathsf{Ran}\}$, which, together with the sets $\mathsf{roleWtn}^{\xi}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2)$ and $\mathsf{lhsWtn}^{\xi}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2)$, is sufficient to decide the existence of a logical difference between $\mathcal{T}_1$ and $\mathcal{T}_2$. However, in practical applications users may require concrete concept inclusions $C \sqsubseteq A$ (or $A \sqsubseteq D$) in $\mathsf{Diff}^{\xi}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2)$ that correspond to a witness A. We note that such *example concept inclusions* can be constructed recursively from triples for which the simulation conditions failed.

4 EXPERIMENTS

To demonstrate the practical applicability of our simulation-based approach for detecting right-hand witnesses in $\mathcal{ELH}^r$-terminologies, we implemented a prototype tool in OCaml that is based on the CEX 2.5 tool [7]. We then conducted a brief experimental evaluation involving large fragments of three versions of SNOMED CT (the first and second international release from 2009 as well as the first international release from 2010) and 119 versions of NCI[8] which appeared between October 2003 and January 2014. The considered fragments of SNOMED CT each contain about 280 000 concepts

[8] More precisely, we first extracted the $\mathcal{ELH}^r$-fragment of the NCI versions by removing up to 8% of the axioms which were not in this fragment.

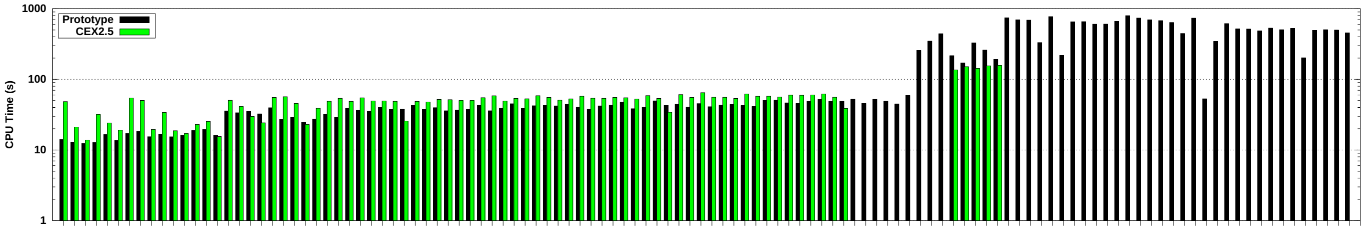

Figure 2. Experimental Results obtained for NCI

names and 62 role names. The aim of our experiments was to compare the performance of our prototype implementation against the CEX 2.5 tool, which can detect logical differences between acyclic terminologies only. We instructed both tools to compute the set $\mathsf{Wtn}^{\mathsf{Ran}}_{\Sigma}(\mathcal{T}_1, \mathcal{T}_2)$ for various versions $\mathcal{T}_1$ and $\mathcal{T}_2$ of SNOMED CT and NCI with $\Sigma = \mathsf{sig}(\mathcal{T}_1) \cap \mathsf{sig}(\mathcal{T}_2)$. All the experiments were conducted on a PC equipped with an Intel Xeon E5-2640 CPU running at 2.50GHz, and all the computation times we report on are the average of three executions.

The results that we have obtained for experiments involving SNOMED CT are shown in Table 1. The first two columns indicate which versions were used as ontologies $\mathcal{T}_1$ and $\mathcal{T}_2$. The next two columns then show the computation times (CPU time) required by CEX 2.5, with column four depicting the computation times if additionally examples illustrating the witnesses were computed. The last two columns then indicate the computation times of our prototype tool. The times required when additionally examples were computed are shown in the last column. One can see that in all the cases our prototype tool required less time to compute difference witnesses (also together with example inclusions) than CEX 2.5.

For each considered version α of NCI, we computed instance witnesses for $\mathcal{T}_1 = \mathrm{NCI}_\alpha$ and $\mathcal{T}_2 = \mathrm{NCI}_{\alpha+1}$, where $\alpha + 1$ denotes the successor version of α, together with corresponding examples. The results that we obtained are depicted in Figure 2. The computations are sorted chronologically along the x-axis according to the publication date of the version NCI_α. Each pair of bars represents the computation times required by our prototype tool and by CEX 2.5, respectively, for one comparison. In the cases where only one bar is shown, the ontology $\mathcal{T}_1 = \mathrm{NCI}_\alpha$ was cyclic and CEX 2.5 could not be used. The values along the y-axis are given in logarithmic scale.

First of all, we observe that, generally speaking, both tools required longer computation times on more recent NCI versions than on older releases, which could be explained by the fact that the size of NCI versions increased with every new release. More precisely, in the comparisons before version 10.03h our prototype tool could typically compute the witnesses and example inclusions faster than CEX 2.5. However, on later versions our new tool then required slightly longer computation times. One can also see that overall it took the longest time to compute witnesses for cyclic versions of NCI.

$\mathcal{T}_1$	$\mathcal{T}_2$	Time (s) - CEX 2.5	with ex.	Time (s) - Prototype	with ex.
SM09a	SM09b	632.06	1839.66	437.28	494.17
SM09b	SM10a	727.60	1190.46	615.84	673.85
SM09b	SM09a	721.28	1042.45	665.40	691.82
SM10a	SM09b	806.68	998.15	563.85	569.22

Table 1. Experimental Results obtained for SNOMED CT

5 CONCLUSION

We have presented a unifying approach to solving the logical difference problem for possibly cyclic $\mathcal{ELH}^r$-terminologies. We have shown that the existence of *backward simulations* in hypergraph representations of terminologies corresponds to the absence of right-hand witnesses (an analogous correspondence exists between forward simulations and left-hand witnesses). We have demonstrated the applicability of the hypergraph approach using a prototype implementation. The experiments show that in most cases our prototype tool outperforms the previous tool, CEX 2.5, for computing the logical difference. Moreover, our prototype tool could be successfully applied on fairly large *cyclic* terminologies, whereas previous approaches only worked for acyclic (or small cyclic) terminologies.

We plan to further improve our prototype implementation. Moreover, extensions of our techniques to general $\mathcal{ELH}^r$-TBoxes and DL-Lite, or even Horn-$\mathcal{SHIQ}$ ontologies could be investigated.

REFERENCES

[1] Franz Baader, Sebastian Brandt, and Carsten Lutz, 'Pushing the $\mathcal{EL}$ envelope further', in *Proceedings of OWLED'08*, (2008).

[2] *The description logic handbook: theory, implementation, and applications*, eds., Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, Cambridge University Press, 2007.

[3] Andreas Ecke, Michel Ludwig, and Dirk Walther, 'The concept difference for $\mathcal{EL}$-terminologies using hypergraphs', in *Proceedings of DChanges'13*. CEUR-WS.org, (2013).

[4] Jennifer Golbeck, Gilberto Fragoso, Frank Hartel, Jim Hendler, Jim Oberthaler, and Bijan Parsia, 'The national cancer institute's thesaurus and ontology', *Web Semantics: Science, Services and Agents on the World Wide Web*, **1**(1), (2003).

[5] The International Health Terminology Standards Development Organisation (IHTSDO), *SNOMED Clinical Terms User Guide*.

[6] Boris Konev, Michel Ludwig, Dirk Walther, and Frank Wolter, 'The logical difference for the lightweight description logic $\mathcal{EL}$', *JAIR*, **44**, 633–708, (2012).

[7] Boris Konev, Michel Ludwig, and Frank Wolter, 'Logical difference computation with CEX2.5', in *Proceedings of IJCAR'12*, pp. 371–377. Springer, (2012).

[8] Boris Konev, Dirk Walther, and Frank Wolter, 'The logical difference problem for description logic terminologies', in *Proceedings of IJCAR'08*, pp. 259–274. Springer, (2008).

[9] Roman Kontchakov, Frank Wolter, and Michael Zakharyaschev, 'Logic-based ontology comparison and module extraction, with an application to DL-Lite', *Artificial Intelligence*, **174**(15), 1093–1141, (October 2010).

[10] Domenico Lembo, Valerio Santarelli, and Domenico Fabio Savo, 'Graph-based ontology classification in OWL 2 QL', in *Proceedings of ESWC 2013*, volume 7882 of *LNCS*, pp. 320–334. Springer, (2013).

[11] Riku Nortje, Arina Britz, and Thomas Meyer, 'Module-theoretic properties of reachability modules for SRIQ', in *Proceedings of DL'13*, pp. 868–884. CEUR-WS.org, (2013).

[12] Boontawee Suntisrivaraporn, *Polytime reasoning support for design and maintenance of large-scale biomedical ontologies*, Ph.D. dissertation, TU Dresden, Germany, 2009.