Highlighting assertional effects of ontology editing activities in OWL

Viktoria Pammer[1,2], Luciano Serafini[3], and Stefanie Lindstaedt[2]

[1] Knowledge Management Institute, TU Graz. Inffeldgasse 21a, 8010 Graz, Austria.
viktoria.pammer@tugraz.at
[2] Know-Center. Inffeldgasse 21a, 8010 Graz, Austria.
slind@know-center.at
[3] Fondazione Bruno Kessler. Via Sommarive, 18, 38100 Trento, Italy.
serafini@fbk.eu

Abstract. Within sufficiently large knowledge bases it is difficult for each contributor to maintain an overview of existing axioms and how they interact with the underlying data. To maintain conceptual consistency however, it is essential that each contributor is able to judge the effects of her actions when manipulating the knowledge base. Assertional effects, exemplary facts which are *lost* or *gained* when removing or adding terminological axioms, give contributors an easy-to-understand means to do so. In this paper we formally define the problem of assertional effects and show that it is decidable for the description logics $\mathcal{SHOIQ}$ and $\mathcal{SROIQ}$ which underlie OWL 1 and OWL 2 respectively. A prototypical implementation which finds approximate solutions accompanies this paper.

1 Introduction

Manual creation and maintenance of formal ontologies requires a significant amount of human effort. This severely impedes the realisation of the Semantic Web. Following success stories of massively user generated content in the vein of Web 2.0, collaboratively maintained description logic knowledge bases may be the way to distribute this effort between many actors. We envision that in the future, contributors will not only add content, as e.g. in Wikipedia[4], or facts, as e.g. in the Semantic Wikipedia envisioned by Krötzsch et al.[1], to a knowledge base but also add terminological axioms to the underlying ontology. Such a scenario is distinguished from traditional ontology engineering scenarios in three main aspects: First, the knowledge engineering process is highly iterative in that a large number of comparatively small revision steps instead of one "ontology creation" phase is to be expected. Second, contributors vary extremely with respect to their (knowledge engineering) expertise. Finally, each single contributor has only limited overview and control of the knowledge base's design and content.

[4] http://www.wikipedia.org/

We propose to capitalize on the existence of meaningful data in such a knowledge base in order to support contributors during ontology editing. In particular, our approach makes the effects of removing or adding terminological axioms to the knowledge base visible in terms of *knowledge lost or gained about data*. This will help contributors to fully understand the results of their actions on the knowledge base. Additionally, the formulation of effects in terms of assertions about instance data is expected to be easily intelligible also for contributors with little knowledge engineering expertise.
The contribution of this paper lies first in motivating such an approach, second in the presentation of a formal characterization of the problem, and third in putting down conditions on the underlying description logic under which the problem is decidable. From this theoretical discussion, a decision algorithm immediately follows. A prototypical implementation called *TeA*[5] accompanies this paper. TeA however does not carry out the full decision procedure but only a pragmatic approximation.

2 Motivation

Consider a knowledge base which contains the facts that "Crete and Kos are islands" (1,2), that "Crete is located in Greece" (3) and that "Crete is located in the Mediterranean Sea" (4). Suppose also that the knowledge base is defined on top of an ontology which defines a set of key concepts such as islands, areas of land vs. areas of water, nations etc. Then suppose that a contributor wants to formalise the concept $Island$ and adds an axiom stating that all islands are located (only) in areas of water (5).

$$Island(crete) \tag{1}$$
$$Island(kos) \tag{2}$$
$$locatedIn(crete, greece) \tag{3}$$
$$locatedIn(crete, mediterranean) \tag{4}$$
$$Island \sqsubseteq \forall locatedIn.WaterArea \tag{5}$$

The most obvious consequence of the additional axiom is that *greece* will then be inferred to be a $WaterArea$, which is surely not intended.

From the logical point of view, the new knowledge base is consistent. However there is still a form of conceptual inconsistency which is due to the fact that the ontology entails a statement inconsistent with the world that the modeller has in mind. Hence, the resulting theory will be represent the point of view of the modeller inadequately or incorrectly. In order to indicate this situation we say that the resulting ontology is *inadequate* or *incorrect*. We point out that this problem is of conceptual nature, in contrast to more formal problems like an unsatisfiable concept or a logically inconsistent knowledge base. Apart from

[5] *TeA*: http://services.know-center.tugraz.at:8080/TeA/Tea.html

inconsistent vocabulary use, sources of conceptual inconsistency also include differing design preferences by contributors, divergent to incompatible underlying views on the domain or simple modelling errors originating in (too) little experience with formal knowledge representation.
Considerations of this kind have led us to study the effects of terminological axioms on the instance data in a knowledge base. In this work, we consider *inferred facts caused by additional terminological axioms* as effects. Where terminological axioms are removed from a knowledge base, *inferences that are lost* and thus "not known anymore" by the knowledge base are considered as effects.

Giving immediate feedback on the effect of ontology edits in terms of concrete individuals gives the contributing users an easy means to review their actions in the light of effects on the whole. This is in line with the realisation that an inherent difficulty in ontology engineering is that such effects are not obvious. This contrasts with the situation in e.g. software engineering where the consequences of one's changes are immediately executable and thus visible (see also [2]). It is precisely this point that a reasoning service computing the effects of axioms on instance data changes. From this perspective, effects in terms of instance data serve as examples of how the terminological axioms will "work" on the knowledge base's data.
Such a reasoning service is therefore especially relevant in the knowledge bases maintained in the spirit of Web 2.0, since (i) frequent ontology edits are expected, (ii) the group of contributors is expected to be heterogeneous in terms of knowledge engineering expertise, but also in terms of views on the knowledge itself and (iii) it follows that such a knowledge base is in danger of becoming chaotic if not each contributor is able to judge the effects of her actions correctly and efficiently.

3 Preliminaries

A vocabulary Σ consists of a set of concept names N_C, a set of role names N_R and a set of individual names N_I. In description logics (DLs), concepts and roles are inductively defined by a set of *constructors* operating on the concept, role and possibly individual names of Σ. Concepts (resp. roles) who are described just by a concept (resp. role) name are also called *primitive* or *named concepts* (resp. roles). The set of complex concepts $\mathcal{C}$ is determined by Σ and a description logic $\mathcal{DL}$. In case these parameters are relevant to the discussion, we write $\mathcal{C}(\Sigma, \mathcal{DL})$. We use A, B to denote primitive concepts, C, D to denote possibly complex concepts, R, S to denote a primitive role and a, b, x, y to denote individuals.

Terminological axioms describe the relation between concepts. An *inclusion axiom* is a specific terminological axiom which of the form $C \sqsubseteq D$, and is often verbalised as "D subsumes C" A set of terminological axioms constitutes a *terminological box*, the so called TBox. In analogy, *role axioms* describe the relation between roles or properties of single roles as for instance symmetry. A set of role axioms constitutes a *role box*, the so called RBox. Terminological and role axioms describe general truths in the domain of discourse. *Assertional axioms* describe

knowledge about individuals. A *concept (resp. role) assertion* for instance is a statement of the form $C(x)$ (resp. $R(x, y)$). These are often verbalised as "x is of type C" and "x is related to y via R" respectively. A set of assertional axioms constitute an *assertional box*, the so called ABox. By a DL *ontology*, a TBox $\mathcal{T}$ and an RBox $\mathcal{R}$ is understood, while by *knowledge base* a TBox, an RBox and an ABox $KB = (\mathcal{T}, \mathcal{R}, \mathcal{A})$ is meant. Assertional axioms are also called *facts*.
An *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of a non-empty set $\Delta^{\mathcal{I}}$, the domain of interpretation, and the interpretation function $\cdot^{\mathcal{I}}$ that assigns to every primitive concept $A \in N_C$ a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, to every primitive role $R \in N_R$ a binary relation $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ and to every individual $x \in N_I$ an element $x^{\mathcal{I}} \in \Delta^{\mathcal{I}}$. An interpretation satisfies a concept inclusion axiom $C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, a concept assertion $C(x)$ iff $x^{\mathcal{I}} \in C^{\mathcal{I}}$ and a role assertion $R(x, y)$ iff $(x^{\mathcal{I}}, y^{\mathcal{I}}) \in R^{\mathcal{I}}$). An ontology (resp. knowledge base) entails an axiom α if and only if all its interpretations satisfy α. This is written as $\mathcal{T} \models \alpha$ resp. $KB \models \alpha$. An interpretation which satisfies an ontology (resp. knowledge base) is also called a *model of the ontology (resp. knowledge base).*

In particular, all discussions in this work target DLs which contain at least $\mathcal{ALC}$, in which complex concepts are constructed as shown in Table 1. Additionally, $\bot$ is used to abbreviate $\neg\top$, $C \sqcup D$ abbreviates $\neg(\neg C \sqcap \neg D)$ and $\exists R.C$ abbreviates $\neg\forall R.\neg C$. In all DLs which include $\mathcal{ALC}$, every TBox $\mathcal{T}$ can be assumed to be in the form $\top \sqsubseteq C_T$ without loss of generalisation: If $\mathcal{T} = \{C_i \sqsubseteq D_i\}$, then $\mathcal{T} = \top \sqsubseteq \bigsqcap_i \neg C_i \sqcup D_i$. Important description logics are $\mathcal{SHOIQ}$ and $\mathcal{SROIQ}$, both of which include $\mathcal{ALC}$. The proposed standard languages for the Semantic Web OWL 1 [3] and OWL 2 [4] are based on $\mathcal{SHOIQ}$ and $\mathcal{SROIQ}$ respectively. For a comprehensive list of features for $\mathcal{SHOIQ}$ we refer to [5], and for $\mathcal{SROIQ}$ we refer to [6] .

Name	Syntax	Semantics
Universal concept	$\top$	$\Delta^{\mathcal{I}}$
Negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
Conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
Universal restriction	$\forall R.C$	$\{w \in \Delta^{\mathcal{I}} \mid \forall v.\ (w, v) \in R^{\mathcal{I}} \rightarrow v \in C^{\mathcal{I}}\}$

Table 1. Syntax and semantics of $\mathcal{ALC}$.

4 Assertional effects of ontology editing activities

By ontology editing activities we understand quite narrowly the addition or removal of terminological axioms to/from an ontology. We call one such activity also an *ontology edit*. In this work we are not concerned with the manipulation of role axioms or facts. The following definition also restricts the meaning of effects to *concept assertions*. Possible extensions with regard to these limitations are part of our ongoing research and discussed in Section 6.

Definition 1. (Assertional effects) *Let* $KB = (\mathcal{T}_0, \mathcal{R}, \mathcal{A})$ *and* $KB_T = (\mathcal{T}_0 \cup \mathcal{T}, \mathcal{R}, \mathcal{A})$. *Let* $\Sigma = (N_C, N_R, N_I)$ *be the vocabulary in which* KB_T *is formulated and let* $\mathcal{DL}$ *be the DL in which the assertional effects shall be formulated.*

- $C(x)$ *such that* $C \in \mathcal{C}(\Sigma, \mathcal{DL})$ *is an assertional effect of* $\mathcal{T}$ *on* KB *iff* $KB \not\models C(x)$ *and* $KB_T \models C(x)$.
- $\mathcal{T}$ *affects an individual* $x \in N_I$ *in* KB *iff an effect* $C(x)$ *of* $\mathcal{T}$ *on* KB *exists.*
- $\mathcal{T}$ *affects a knowledge base* KB *iff there exists an individual* $x \in N_I$ *such that* $\mathcal{T}$ *affects* x *in* KB.

Since this work is only concerned with assertional effects of ontology edits, we sometimes omit the "assertional" and just speak of effects.
As preconditions we assume that KB_T is consistent and that N_I is not empty, i.e. the knowledge base knows at least about one individual. Individuals can occur in $\mathcal{A}$ or, if the DL allows for nominals as $\mathcal{SHOIQ}$ and $\mathcal{SROIQ}$ do, also in $\mathcal{T}$.

The above definition can be applied to both ontology editing activities. If $\mathcal{T}$ is added to KB, then the effects of $\mathcal{T}$ on KB represent knowledge about individuals which is gained. If $\mathcal{T}$ is removed from KB_T, then the effects of $\mathcal{T}$ on KB represent the knowledge about individuals which is lost.

4.1 Deciding the Existence of Assertional Effects

Since the set $\mathcal{C}(\Sigma, \mathcal{DL})$ is in general infinite for DLs equally or more expressive than $\mathcal{ALC}$, we consider at first the decidability of the general question whether a particular TBox $\mathcal{T}$ affects a particular knowledge base KB. In order to do so, we first define the notion of *reachability* of a concept C from an individual x in a knowledge base KB.
Let R^- denotes an inverse role such that $(R^-)^{\mathcal{I}} = \{(y, x) | (x, y) \in R^{\mathcal{I}}\}$, and let $N_R^- = \{R | R \in N_R \text{ or } R^- \in N_R\}$ in description logics which include inverse roles.

Definition 2. $w_0 R_1 w_1 \ldots R_n w_n$ is called a path in an interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ *if and only if for* $i = 1 \ldots n$ *it holds that* $w_i \in \Delta^{\mathcal{I}}$, *and* $(w_{i-1}, w_i) \in R_i^{\mathcal{I}}$ *and* $R_i \in N_R^-$.

Definition 3. A concept C is reachable from $x \in N_I$ w.r.t. KB *iff there is a model* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ *of* KB *in which* $x^{\mathcal{I}} = w_0$, *and a path* $w_0 R_1 w_1 \ldots R_n w_n$ *exists in* $\mathcal{I}$ *such that* $w_n \in C^{\mathcal{I}}$.

In other words, C is reachable from x in KB iff either $C(x)$ or $\exists R_1 \ldots R_n . C(x)$ for $n > 0$ is satisfiable w.r.t. KB.
The definition of reachability is motivated by the fact that it can be shown that reachability equals the existence of assertional effects. First, this is shown under the condition that the DL in question is decidable under a tableaux decision procedure. Later, we will show a small generalisation.

Theorem 1 *In description logics decidable under a tableaux decision procedure,* $\mathcal{T} = \{\top \sqsubseteq C_T\}$ *affects KB iff an individual* x *exists in KB such that* $\neg C_T$ *is reachable from* x *in KB.*

For both $\mathcal{SHOIQ}$ and $\mathcal{SROIQ}$ tableaux decision procedures exist. For exact descriptions of tableaux algorithms we refer the reader to [5] for $\mathcal{SHOIQ}$ and to [6] for $\mathcal{SROIQ}$. Nonetheless, we review some basic notions related to tableaux decision procedures before delving into the proof for Theorem 1.

A *completion graph* for a knowledge base KB formulated in the vocabulary Σ and the description logic $\mathcal{DL}$ is a labelled directed graph $G = (V, E, \mathcal{L}, \not\doteq)$ where each node $x \in V$ is labelled with a set $\mathcal{L}(x) \subseteq \mathcal{C}(\Sigma, \mathcal{DL})$ and each edge $(x, y) \in E$ is labelled with a set of role names $\mathcal{L}(x, y) \subseteq N_R^-$. A set of *completion rules* are used to manipulate the underlying completion graph(s).
A completion graph contains a *clash* iff a label $\mathcal{L}(x)$ contains either $\bot$ or both A and $\neg A$. A completion graph which does not contain a clash is called *open*, while a a completion graph which contains a clash is called *closed*. A completion graph to which no more completion rules apply is called *complete*.
More specifically we make use of the following connections between consistency, completion graphs and models, which hold whenever a tableaux algorithm has been shown to be a decision procedure for a DL language.

- **If a knowledge base is consistent, then an open and complete completion graph can be constructed.** This is the basis of the completeness property of tableaux decision algorithms.
- **Every open and complete completion graph can be translated into a model.** This is the basis of the correctness property of tableaux decision algorithms. The relevant part of this translation is the following: If a completion graph $G = (V, E, \mathcal{L}, \not\doteq)$ is translated into a model $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, then every node $x \in V$ corresponds to at least one node $w \in \Delta^{\mathcal{I}}$ such that for all $C \in \mathcal{C}(\Sigma, \mathcal{DL})$, $C \in \mathcal{L}(x)$ if and only if $w \in C^{\mathcal{I}}$, and every edge $(x, y) \in E$ corresponds to at least one relation $(v, w) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ such that for all $R \in N_R^-$, $R \in \mathcal{L}(x, y)$ if and only if $(v, w) \in R^{\mathcal{I}}$.
- **All nodes in a completion graph for a knowledge base** KB **are either an individual or connected to an individual iff** KB **contains at least one individual.** This property follows from the procedure of tableaux-based algorithms, which start with a set of initial nodes consisting of individual nodes (individuals or nominals) and create only new nodes which are connected to an existing node. If KB does not contain any individual however, tableaux-based algorithms typically start with an "invented" initial single node labelled with C_T given that $\mathcal{T} = \top \sqsubseteq C_T$. Remember that the existence of individuals was assumed as a precondition.

In the following we say that a concept C can be consistently added to the label of a node $w \in V$ of a completion graph $G = (V, E, \mathcal{L}, \not\doteq)$ iff G can be completed into an open and complete graph using the completion rules after C is added to the label of w.

Proof. Let $KB = (\mathcal{T}_0, \mathcal{A})$ and $KB_T = (\mathcal{T}_0 \cup \mathcal{T}, \mathcal{A})$.
$\Leftarrow$ **If $\neg C_T$ is reachable from an individual x in KB, $\mathcal{T}$ affects x in KB**
If $\neg C_T$ is reachable from an individual x, there is an $n \in \mathbb{N}_0$ such that if $n = 0$ then $\neg C_T(x)$ or if $n > 0$ then $(\exists R_1 \dots R_n.\neg C_T)(x)$ is satisfiable w.r.t. KB. If $n = 0$ this means that $KB \not\models C_T(x)$ or if $n > 0$ this means that $KB \not\models (\forall R_1 \dots R_n.C_T)(x)$. On the other hand, $\top \sqsubseteq C_T \models \top \sqsubseteq \forall R_1 \dots R_n.C_T$ is trivially true, and therefore $KB_T \models C_T(x)$ and $KB_T \models (\forall R_1 \dots R_n.C_T)(x)$.
$\Rightarrow$ **If $\mathcal{T}$ affects KB, then an individual x exists such that $\neg C_T$ is reachable from x in KB**
Proof by contradiction, i.e. it is assumed that an effect exists but that $\neg C_T$ is not reachable from x in KB.
Let $D(x)$ be one of the possibly many effects of $\mathcal{T}$ on KB.
Since $\neg D(x)$ is consistent w.r.t. KB, an open and complete completion graph $G = (V, E, \mathcal{L}, \neq)$ for KB can be constructed such that $\neg D \in \mathcal{L}(x)$.
$\neg D(x)$ is inconsistent with the extended knowledge base KB_T. Therefore the following procedure, extending the open and complete graph G leads to only closed completion graphs: Add C_T to the label of a node in V. Follow the completion rules, and ensure that nodes newly created in the process are also labelled with C_T. Repeat for all nodes in G until for one node w_C adding C_T to $\mathcal{L}(w_C)$ leads to only closed completion graphs.
Then however, $\neg C_T$ can be consistently added to $\mathcal{L}(w_C)$.
Since all nodes in a completion graph either are an individual node or connected to one, there is then an individual $y \in V$ from which a path to w_C can be constructed. Call this path $yR_1w_1 \dots R_nw_C$. If $n = 0$, then $y = w_C$.
G can be translated into a model $\mathcal{I}$ such that $w_C \in (\neg C_T)^{\mathcal{I}}$, and $(y, w_1) \in R_1^{\mathcal{I}}$, $(w_1, w_2) \in R_2^{\mathcal{I}}$, …, $(w_{n-1}, w_C) \in R_n^{\mathcal{I}}$. Then, $\neg C_T$ is reachable in KB from y.

As by-product from the equivalence between axiom effects and reachability the following corollary can be derived.

Corollary 1 *If $\mathcal{T} = \{\top \sqsubseteq C_T\}$ affects KB, then an effect $C(x)$ exists such that $C \doteq \forall R_1 \dots R_n.C_T$ and $R_i \in \Sigma, i = 1 \dots n$. If $n = 0$, this corresponds to $C = C_T$. n is bounded by the maximal number of nodes in completion graphs for the corresponding description logic.*

Then, the following theorem about decidability follows immediately:

Theorem 2 *The existence of assertional effects of $\mathcal{T}$ on KB can be decided in all logics decidable under tableaux algorithms.*

Some interesting observations follow from the these results: First, in order to express such effects, DLs which contain at least $\mathcal{ALC}$, i.e. which include negation over complex concepts and qualified universal/existential quantification, are required. Second, if an effect exists, then not all effects are necessarily of the form $C(x)$ with $C \doteq \forall R_1 \dots R_n.C_T$. As a simple example consider extending the knowledge base $KB = \{R(a, b)\}$ is with the TBox $\mathcal{T} = \{\top \sqsubseteq \forall R.A\}$. In this case, the effect $(\forall R.A)(a)$ will be found if looking for effects of the above-mentioned

pattern, but clearly also $A(b)$ is an effect. Third, although bounded, n can be quite high: In $\mathcal{SHOIQ}$, n is bounded double-exponentially with the size of the closure of $\mathcal{T}_0$ (the smallest set containing all subconcepts of $\mathcal{T}_0$ and closed under negation) and the number of roles and inverses occurring in the input [5, 7].

4.2 Generalisation to DLs with the connected model property

In description logics which additionally provide role union and the reflexive-transitive closure (Kleene operator *), "C is reachable from x in KB" can also be expressed as $\exists(\bigsqcup_{R_i, R_i^- \in N_R} R_i)^*.C(x)$. This led to the question of whether Theorem 1, which states equivalence between assertional effects and reachability, can be generalised to require a more general property of the underlying description logic than being decidable under a tableaux decision procedure. Indeed, it can be shown that only the *connected model property* is required:

Definition 4. Connected model *A model $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ is connected if and only if for every $w \in \Delta^{\mathcal{I}}$ there is an element $x \in N_I$, $x^{\mathcal{I}} = w_o$ such that there is a path $w_0 R_1 w_1 \dots R_n w$ in $\mathcal{I}$.*

A logic is said to have the *connected model property* if every satisfiable concept or consistent knowledge base has a connected model. Since tree and forest models are connected models, all logics which enjoy the tree (forest) model property, also have the connected model property.

Theorem 3 *In description logics with the connected model property $\mathcal{T} = \{\top \sqsubseteq C_T\}$ affects KB iff an individual x exists in KB such that $\neg C_T$ is reachable from x in KB.*

Proof. Theorem 3 Let $KB = (\mathcal{T}_0, \mathcal{A})$ and $KB_T = (\mathcal{T}_0 \cup \mathcal{T}, \mathcal{A})$.
$\Leftarrow$ **If $\neg C_T$ is reachable from an individual x in KB, $\mathcal{T}$ affects x in KB**
This direction is the same as in the proof for the tableaux-based Theorem 1.
$\Rightarrow$ **If $\mathcal{T}$ affects KB, then an individual x exists such that $\neg C_T$ is reachable from x in KB**
Proof by contradiction, i.e. it is assumed that an effect exists but that $\neg C_T$ is not reachable from x in KB.
Let $D(x)$ be one of the possibly many effects of $\mathcal{T}$ on KB.
Then there is a connected model $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ of $\neg D(x)$ w.r.t.
$\mathcal{I}$ is not a model however of the extended knowledge base KB_T, since $\neg D(x)$ is inconsistent w.r.t. KB_T.
Therefore, there is an element $w \in \Delta^{\mathcal{I}}$ such that $w \notin (C_T)^{\mathcal{I}}$. Otherwise, $\Delta^{\mathcal{I}} = (C_T)^{\mathcal{I}}$ and $\mathcal{I}$ would also be a model of KB_T.
Since $w \notin (C_T)^{\mathcal{I}}$, it holds that $w \in (\neg C_T)^{\mathcal{I}}$.
Because of the connected model property, there is then an individual $y \in N_I$ and $y^{\mathcal{I}} = w_0$ such that there is a path $w_0 R_1 w_1 \dots R_n w_C$. If $n = 0$ this means that $w_0 = w_C$. Then, $\neg C_T$ is reachable from the individual y, which contradicts the assumption that $\neg C_T$ is not reachable from any individual in the knowledge base.

Therefore, the problem of deciding whether $\mathcal{T}$ affects KB can be posed as consistency checks of the form $\exists(\bigsqcup_{R_i,R_i^- \in N_R} R_i)^*.\neg C_T(x)$ for all $x \in N_I$ in DLs which have the connected model property and the required concept and role constructors. Such a reformulation has the advantage of posing the original problem as a standard reasoning problem. Naturally, this reformulation only makes sense if the resulting logic is also decidable. A logic for which all these requirements hold is for instance $\mathcal{ALCCQI}b^+_{reg}$, which has been shown to be decidable [8]. Unfortunately no reasoners exist to date for this DL to the best of our knowledge.

4.3 TeA: Prototypical Implementation

The TeA prototype[6] is, from a user's point of view, a website where users can upload a knowledge base, add terminological axioms and see the resulting assertional effects. The TeA prototype accepts OWL knowledge bases and OWL axioms as input. The current prototype can not be directly used to manipulate knowledge bases in any way but simply demonstrates the computation of assertional effects. The TeA backend functionality will be integrated at some point into the MoKi (a wiki environment for modelling) however, and there will also support all editing activities, i.e. both deletion and addition of axioms. As follows from Corollary 1, the depth n of an effect $(\forall R_1 \dots R_n.C_T)(x)$ could be very large. In TeA a pragmatic approximation was implemented which limits n. Hence, all assertional effects delivered by TeA are correct but TeA is not guaranteed to find effects.

5 Related work

5.1 Conservative extensions in DL

Both conceptually and technically, conservative extensions in description logics are a closely related topic. In short, deciding conservativity for two TBoxes $\mathcal{T}_0$ and $\mathcal{T}$ means finding out whether $\mathcal{T}_0 \cup \mathcal{T}$ entails any inclusion axioms expressible in a given vocabulary and a given DL that are not entailed by $\mathcal{T}_0$ alone [9, 10]. It can easily be shown that non-conservativity of $\mathcal{T}_0 \cup \mathcal{T}$ with respect to $\mathcal{T}_0$ is a precondition for the existence of TBox effects of $\mathcal{T}$ on $KB = (\mathcal{T}_0, \mathcal{R}, \mathcal{A})$: By inventing an ABox $\mathcal{A} = \{\top(x)\}$, any inclusion axiom entailed by $\mathcal{T}_0 \cup \mathcal{T}$ but not by $\mathcal{T}_0$ alone produces an effect on x.
However, non-conservativity is not a guarantee for the existence of effects. To illustrate the latter, consider the following knowledge base $KB = (\mathcal{T}_0, \mathcal{R}, \mathcal{A})$:

$$\begin{aligned} &A \sqsubseteq \forall R.A \\ &A \sqsubseteq \exists R.A \\ &A(x) \end{aligned} \tag{6}$$

which is extended with $\mathcal{T}$:

$$\top \sqsubseteq A \tag{7}$$

[6] *TeA*: http://services.know-center.tugraz.at:8080/TeA/Tea.html

Obviously, $\mathcal{T}_0 \cup \mathcal{T}$ is not a conservative extension of $\mathcal{T}_0$ w.r.t. $\Sigma = \{A, R, x\}$ and the description logic $\mathcal{ALC}$. There are however, no effects on the individual x, since KB already entails all types that can be constructed from the vocabulary $\{A, R\}$ for x in $\mathcal{ALC}$.
Complexity results for deciding conservativity therefore give a *lower bound* on the complexity of deciding effects according to the deductive definition.

Depending on the choice of vocabulary, conservativity can be reduced to subsumption if the full vocabulary of KB is considered [9]. Interestingly, the problem becomes harder if the vocabulary under consideration is a subset of the vocabulary used by KB. Then, conservativity is decidable up to $\mathcal{ALCQI}$ [10].
Using these results from conservativity now opens up the possibility to extend the notion of assertional TBox effects as considered so far. Remember that assertional TBox effects on a knowledge base have been defined only for the case where Σ is the vocabulary in which KB_T is formulated (Definition 4). If a contributor is interested in effects in terms of a smaller vocabulary $\Sigma' \subset \Sigma$, the following procedure can be taken to circumvent this small restriction: Given is the knowledge base $KB = (\mathcal{T}_0, \mathcal{R}, \mathcal{A})$ which shall be extended with the TBox $\mathcal{T}$. Let $KB_T = (\mathcal{T}_0 \cup \mathcal{T}, \mathcal{R}, \mathcal{A})$ and let Σ be the vocabulary in which KB_T is formulated. Furthermore we assume that $\Sigma' \subset \Sigma$ is the vocabulary and $\mathcal{DL}$ the description logic in which the assertional effects on KB shall be formulated. Then, in a first step it must be decided whether $\mathcal{T}_0 \cup \mathcal{T}$ is a conservative extension of $\mathcal{T}_0$ w.r.t. Σ' and $\mathcal{DL}$ can be decided. If $\mathcal{T}_0 \cup \mathcal{T}$ is not a conservative extension of $\mathcal{T}_0$ w.r.t. Σ' and $\mathcal{DL}$, then a witness concept C'_T such that $\neg C'_T$ is satisfiable w.r.t. $\mathcal{T}_0$ but unsatisfiable w.r.t. $\mathcal{T}_0 \cup \mathcal{T}$ exists. A decision procedure for conservativity such as e.g. in [10] outputs such a witness concept[7]. Given this C'_T, the question of whether and which assertional effects of $\mathcal{T}$ on KB exist can be reformulated to the question whether $\top \sqsubseteq C'_T$ affects KB, under the condition that $\mathcal{DL}$ contains at least $\mathcal{ALC}$.

Conceptually, we stress the difference in underlying motivation between conservative extensions and assertional TBox effects on a knowledge base. Comparing TBoxes for differences is a general approach to support the frequent task of extending or refining an ontology. The rationale behind focusing on effects in terms of instance data is directed towards ontology edits in a specific ontology application scenario, namely where the ontology describes data in a knowledge base. In this scenario it is important that ontology and data are well aligned with each other in order to maintain conceptual consistency. Second, expressing effects of terminological axioms (general truths in a domain) in terms of concrete facts illustrates them in an easily understandable way[8]. This gives users an opportunity to double-check on whether the effectuated changes were actually "meant this way".

[7] Note that this particular decision procedure would actually output the negation of C'_T.

[8] Compare also [2], in which the creation of concept definitions from exemplary individuals is being discussed for exactly the same reason. It often seems to be helpful to think in concrete terms when formulating abstractions.

5.2 Reasoning services for ontology engineering

In continuance of the idea underlying conservativity, Kontchakov et al [11] study the differences between DL-Lite TBoxes. Especially, the authors study query-differences over arbitrary ABoxes. Then the set of query-differences is either empty (the two TBoxes do not differ in terms of queries given a specific vocabulary at all) or infinite (infinitely many ABoxes exist after all). Naturally it would be interesting to consider effects in terms of queries over a specific ABox, which we have not done so far. We speculate that the results for such a problem formulation will be similar than the comparison of deductively defined effects with conservative extension, namely that it is at least as hard as deciding the existence of query-differences, and not immediately clear how the existence of query-differences then can be decided for a given ABox.

On a more general note, ontology editors like Swoop [12] and Protégé [13] display *inferred types* for all individuals and concepts in the loaded knowledge base. Typically, such inferred types involve only primitive concepts. Additionally, the dynamic aspect of the ontology edits is not considered in that inferred types are shown for a complete knowledge base, and no relation is automatically made to the most recent activities.

Under the names of ontology debugging and repair, reasoning services which pinpoint axioms responsible for unsatisfiable concepts and suggest ways to repair the ontology have been researched, see e.g. [14, 15]. Inference explanation, e.g. [16, 17], is based on the same theoretical foundation, but with a different focus in application, as is already suggested through the naming. Both ontology repair and inference explanation start with an identified problem, whereas the study of assertional effects seeks to make potential problems (effects) visible. In this sense, repair and explanation are complementary reasoning services useful *after having identified a problematic effect.*

6 Discussion and Outlook

Informative Effects Consider as slightly extended example a knowledge base which contains diverse facts such as "Crete is an Island", "The Mediterranean is a Sea" and "Sophocles is a Greek" (8) . It is overlaid with an ontology that defines relevant key concepts such as areas of land vs. areas of water, nations etc, and that an island is a land-area, while a sea is a water-area, defined as subsumption. One contributor decides to increase the accuracy of the ontology, and adds a disjointness axiom for land- and water-areas (9), and an existential restriction which states that each island must be located in some water-area (10).

$$\begin{aligned} &Greek(sophocles) \\ &Island(crete) \qquad (8) \\ &Sea(mediterranean) \end{aligned}$$

$$LandArea \sqcap WaterArea \sqsubseteq \bot \qquad (9)$$

$$Island \sqsubseteq \exists locatedIn.WaterArea \qquad (10)$$

Then, depending on which kinds of facts "count", effects of the disjointness axioms are $\neg WaterArea(crete)$, $\neg LandArea(mediterranean)$ but also $(\neg WaterArea \sqcup \neg LandArea)(sophocles)$. Additionally, the existential restriction $(\exists locatedIn.WaterArea)(crete)$ but also $(\neg Island \sqcup \exists locatedIn.WaterArea)(sophocles)$ are newly entailed facts.

As a personal rating, we would consider $\neg WaterArea(crete)$, $\neg LandArea(mediterranean)$ and $(\exists locatedIn.WaterArea)(crete)$ to be more informative than for instance $(\neg WaterArea \sqcup \neg LandArea)(sophocles)$ or $(\neg Island \sqcup \exists locatedIn.WaterArea)(sophocles)$. This short list demonstrates that some assertional effects are more informative than others, and that this quality does not directly depend on whether complex concepts are involved or not. It is to date not completely clear which characteristics determine how informative an assertional effect is. This shall be captured by a user study as part of ongoing research.

Exemplary effects A critical issue concerning assertional effects in the envisioned scenario concerns the quantity of data to be dealt with. Intuitively, if a knowledge base contains data about a million song-titles and a new axiom stating that "every song has an author" is added which, it is clearly not desirable to see a million effects of the form "The song XY has an author". Instead, assertional effects should be expressed using *exemplary* individuals only. A lead into that direction could be given by techniques such as ABox summarization [18]. The authors exploit the observation that similar individuals are related in similar ways to other individuals. For instance, songs have titles, belong to albums and have maybe been in some charts for a given period of time. Songs do not however have parents or children as do human persons. Thus, individuals about which *similar assertions* exist in a knowledge base can be grouped together. The main issues which need to be studied when applying this ABox summarization to the computation of exemplary assertional effects are that (i) ABox summarization has been defined for $\mathcal{SHIN}$ only in [18] and (ii) the summary ABox does not preserve consistency, i.e. it is possible that the summary ABox is inconsistent while the original ABox is consistent w.r.t. a knowledge base. Apart from these theoretical issues however, the benefit clearly lies not only in enabling an improved presentation to the user but also in increasing the computational performance (though not the computational complexity class).

Extending the definition of assertional effects The definition of assertional TBox effects on a knowledge base (Definition 1) can be extended into a variety of directions. One such extension, namely the possibility to restrict the vocabulary under consideration to a subset of the vocabulary Σ in which KB_T is formulated, has already been discussed in Section 5 in relation with conservative extensions. Another obvious but much more simple extension is to consider also role assertions $R(x, y)$ as assertional effects. This is trivial in DL languages in which the set of roles which can occur in role assertions is finite, as is the case in both $\mathcal{SHOIQ}$ and $\mathcal{SROIQ}$ and thus also in OWL 1 and OWL 2. Then, for every $R \in N_R^{ext}$ and every pair (x, y), $x, y \in N_I$, it simply needs to be checked whether $KB \not\models R(x, y)$

and $KB_T \models R(x, y)$. In $\mathcal{SHOIQ}$, $N_R^{ext} = N_R^- = \{R | R \in N_R \text{ or } R^- \in N_R\}$. For $\mathcal{SROIQ}$, N_R^{ext} must be extended to be closed under negation. We note that unless nominals occur in $\mathcal{T}$, terminological axioms can not cause the gain or loss of role assertion axioms. Furthermore, also equality or inequality assertions between individuals could be considered as effects. Such effects may occur when $\mathcal{T}$ contains nominals or number restrictions.
Finally, the notion of assertional effects could also be extended to consider assertional effects of the addition or removal of role axioms.

7 Conclusion

We anticipate the spreading of collaborative description logic knowledge bases in which users contribute not only by adding textual or multimedia content but also facts about individuals or other axioms. Our approach to ensure conceptual consistency of such knowledge bases is to highlight the consequences of terminological axioms in terms of instance data, namely as facts which are lost or gained. Contributors to a DL knowledge base benefit from this approach in two different ways. First, such assertional effects give concrete examples of the general knowledge (terminological axioms) which was removed or added. Second, assertional effects take into account the complex interactions between all the terminological, role and assertional axioms in the knowledge base.
Based on a formal definition of assertional TBox effects on a knowledge base, we have derived conditions for decidability. In particular, the problem of deciding the existence of assertional effects has been shown to be equivalent to deciding the reachability of a given concept in a knowledge base if the underlying DL has the connected model property. Additionally, we considered only DL languages at least as expressive as $\mathcal{ALC}$ in our discussion. Consequently, our results hold both for $\mathcal{SHOIQ}$ and $\mathcal{SROIQ}$, the DL languages underlying OWL 1 and 2 respectively.
In a detailed discussion we pointed to wider-reaching conceptual issues such as which assertional effects are informative to knowledge base contributors and how assertional effects can be presented to contributors in a meaningful way in the presence of a large quantity of data in the knowledge base. These issues are investigated in ongoing and future research.

Acknowledgements

Some of the authors are supported by APOSDLE (www.aposdle.org), which is partially funded under the FP6 of the European Commission within the IST work program 2004 (FP6-IST-2004-027023). The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

References

1. Krötzsch, M., Vrandecic, D., Vlkel, M., Haller, H., Studer, R.: Semantic wikipedia. Journal of Web Semantics **5** (Sept. 2007) 251–261
2. Lutz, C., Baader, F., Franconi, E., Lembo, D., Möller, R., Rosati, R., Sattler, U., Suntisrivaraporn, B., Tessaris, S.: Reasoning support for ontology design. In: OWL: Experiences and Directions 2006. (2006)
3. OWL Web Ontology Language Reference: http://www.w3.org/tr/owl-ref/, last visited 2008-11-28
4. OWL 2 Web Ontology Language Quick Reference Guide: http://www.w3.org/tr/owl2-quick-reference/, last visited 2009-07-15
5. Horrocks, I., Sattler, U.: A tableau decision procedure for $\mathcal{SHOIQ}$. J. Autom. Reason. **39**(3) (2007) 249–276
6. Horrocks, I., Kutz, O., Sattler, U.: The even more irresistible $\mathcal{SROIQ}$. In: Proc. of the 10th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR 2006), AAAI Press (2006) 57–67
7. Horrocks, I., Sattler, U., Tobies, S.: Reasoning with individuals for the description logic $\mathcal{SHIQ}$. In: CADE-17: Proceedings of the 17th International Conference on Automated Deduction, London, UK, Springer-Verlag (2000) 482–496
8. Ortiz, M.: An automata-based algorithm for description logics around $\mathcal{SRIQ}$. In: Proceedings of the Fourth Latin American Workshop on Logic/Languages, Algorithms and Non-Monotonic Reasoning 2008 (LANMR'08). Volume 408., Puebla,Mexico, CEUR Workshop Proceedings (October 2008)
9. Ghilardi, S., Lutz, C., Wolter, F.: Did I damage my ontology? A case for conservative extensions in description logics. In: Proceedings of KR2006: the 20th International Conference on Principles of Knowledge Representation and Reasoning, 2006, AAAI Press (2006) 187–197
10. Lutz, C., Walther, D., Wolter, F.: Conservative extensions in expressive description logics. In: Proceedings of the Twentieth International Joint Conference on Artificial Intelligence IJCAI-07, AAAI Press (2007)
11. Kontchakov, R., Wolter, F., Zakharyaschev, M.: Can you tell the difference between dl-lite ontologies? In: KR 2008: Eleventh International Conference on Principles of Knowledge Representation and Reasoning. (2008)
12. Swoop: http://code.google.com/p/swoop/. Last visited 2008-11-28
13. Protégé: http://protege.stanford.edu/. Last visited 2008-11-28
14. Kalyanpur, A., Parsia, B., Sirin, E., Hendler, J.: Debugging unsatisfiable classes in OWL ontologies. Journal of Web Semantics **3**(4) (2005) 268–293
15. Schlobach, S., Huang, Z., Cornet, R., Harmelen, F.: Debugging incoherent terminologies. J. Autom. Reason. **39**(3) (2007) 317–349
16. Kalyanpur, A., Parsia, B., Horridge, M., Sirin, E.: Finding all justifications of OWL DL entailments. In: Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007). Volume 4825 of LNCS., Springer Verlag (November 2007) 267–280
17. Horridge, M., Parsia, B., Sattler, U.: Laconic and precise justifications in OWL. In: International Semantic Web Conference. Volume 5318 of Lecture Notes in Computer Science., Springer (2008) 323–338
18. Fokoue, A., Kershenbaum, A., Ma, L., Schonberg, E., Srinivas, K.: The summary abox: Cutting ontologies down to size. In: International Semantic Web Conference. Volume 4273 of Lecture Notes in Computer Science., Springer (2006) 343–356