Adaptive $\mathcal{ALE}$-TBox for Extending Terminological Knowledge

Ekaterina Ovchinnikova[1] and Kai-Uwe Kühnberger[2]

[1] University of Tübingen, Seminar für Sprachwissenschaft
e.ovchinnikova@gmail.com
[2] University of Osnabrück, Institute of Cognitive Science
kkuehnbe@uos.de

Abstract. Ontologies are usually considered as static data structures representing conceptual knowledge of humans. For certain types of applications it would be desirable to develop an algorithmic adaptation process that allows dynamic modifications of the ontology in the case new information is available. Dynamic updates can generate conflicts between old and new information resulting in inconsistencies. We propose an algorithm that can model the adaptation processes for conflicting and non-conflicting updates defined on $\mathcal{ALE}$-TBoxes.

1 Introduction

In artificial intelligence, there is an increasing interest in examining ontologies for a variety of applications in knowledge-based systems. A subsumption relation defined on concepts together with additional (optional) relations is usually considered as the core of every ontology. Several formalisms were proposed to represent ontologies. Probably the currently most important markup language for ontology design is the *web ontology language*[3] (*OWL*) in its three different versions: *OWL light*, *OWL DL*, and *OWL full*. Besides the markup specifications of representation languages, a related branch of research is concerned with logical characterizations of ontology representation formalisms using certain subsystems of predicate logic, so-called Description Logics (DL) [3]. It is well-known that *OWL* standards can be characterized by description logics.

The acquisition of ontological knowledge is one of the most important steps in order to develop new and intelligent web applications [4]. Because hand-coded ontologies of a certain size are usually tedious, expensive, and time-consuming to develop, there is the need for a (semi-)automatic extraction of ontological knowledge from given data. Furthermore rapid changes concerning the information theoretic background and the dramatic increase of available information motivates (semi-)automatic and consistent adaptation processes changing existing ontologies, if new data requires a modification. The aim of the present paper is to develop a theory of adapting terminological knowledge by extending the terminology with additional concepts and relations for conflicting and non-conflicting updates.

[3] See the documentation at http://www.w3.org/TR/owl-features/.

2 Related Work

There is a variety of logical approaches in order to model inconsistencies. Examples are default logic, paraconsistent logic, or many-valued logic. Similarly several ontology extension mechanisms on the basis of classical description logics were proposed. This is quite often realized by the introduction of new construction methods like planning systems [2] or belief-revision processes [8]. The disadvantage is that standard DL-reasoners cannot be used easily for these extensions. Another tradition tries to model inconsistencies without extending the underlying DL: For example, in [9] a witness concept is demonstrating an occurring inconsistency, in order to characterize non-conservative extensions. Unfortunately this does not solve the problem, but just shows where the problem is. Another example of approaches not extending DL is based on learning techniques. Inductive logic programming is used in [7], where a new definition for a problematic concept is generated on the basis of positive examples. In this case as well as in cases where statistical learning techniques are used [6], information is lost due to the fact that the original definition is fully ignored.

We propose a solution in which the ontology remains formalized in DL and new information is adapted to the terminological knowledge automatically keeping the changes as minimal as possible.

3 Description Logics

DL provide a logical basis for representing ontologies [3]. In this paper, we consider mainly $\mathcal{ALE}$ and subsidiary $\mathcal{ALC}$, two relatively weak DLs.

A *TBox*[4] (terminological box) is a finite set of axioms of the form $C \equiv D$ (equalities) or $C \sqsubseteq D$ (inclusions) where C, D (called *concept descriptions*) are specified as follows in the $\mathcal{ALE}$-DL (A stands for atomic concept, R stands for role name): $C \rightarrow \top \mid \bot \mid A \mid \neg A \mid \forall R.C \mid \exists R.C \mid C_1 \sqcap C_2$. $\mathcal{ALC}$-DL additionally introduces the concept descriptions $\neg C$ and $C_1 \sqcup C_2$.

Concept descriptions are interpreted model-theoretically (see [3] for details). An interpretation $\mathcal{I}$ is called a *model* of a TBox $\mathcal{T}$ iff for all $C \sqsubseteq D \in \mathcal{T}$: $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ and for all $C \equiv D \in \mathcal{T}$: $C^{\mathcal{I}} = D^{\mathcal{I}}$. D *subsumes* C *towards* $\mathcal{T}$ ($\mathcal{T} \models C \sqsubseteq D$) iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for every model $\mathcal{I}$ of $\mathcal{T}$. C is *satisfiable towards* $\mathcal{T}$ if there is a model $\mathcal{I}$ of $\mathcal{T}$ such that $C^{\mathcal{I}}$ is nonempty; otherwise C is *unsatisfiable* and $\mathcal{T} \models C \sqsubseteq \bot$. Subsumption algorithms (see [3]) are implemented in several reasoning systems[5].

The symbol $\doteq$ denotes the syntactical equality of concept descriptions. Atomic concepts occurring on the left side of an equality are called *defined*. The set $ax(\mathcal{T})$ collects *axiomatized* concepts occurring on the left side of an axiom in $\mathcal{T}$. *Unfolding* an acyclic $\mathcal{T}$ or a concept description towards $\mathcal{T}$ consists in a syntactical replacement of the defined concepts by their definitions. In order to make $\mathcal{T}$ *unfoldable* all inclusions must be replaced with equalities as follows: $C \sqsubseteq D \rightarrow C \equiv D \sqcap C^*$. A TBox $\mathcal{T}$ is *inconsistent* iff $\exists A \in ax(\mathcal{T}) : \mathcal{T} \models A \sqsubseteq \bot$.

[4] We restrict out attention to the TBox leaving the ABox for further investigations.
[5] Some DL reasoners are listed at `www.cs.man.ac.uk/~sattler/reasoners.html`.

4 Automatic Adaptation to the New Axiom: Solutions

Informally, the *adaptation* of a TBox $\mathcal{T}$ to a new axiom for a concept X is a minimal modification of $\mathcal{T}$ such that X becomes satisfiable towards $\mathcal{T}$. By modification we mean the sequence of adding and deleting axioms in $\mathcal{T}$. Since the notion of minimal modification in general is rather vague, we suggest a constructive definition of the adaptation procedure introduced in Sec. 5.

A conflict between a new axiom $X \sqsubseteq Y$ and the original TBox $\mathcal{T}$ occurs if there exist a concept $C \in ax(\mathcal{T})$ such that $\mathcal{T} \models Y \sqsubseteq C$ and definition D of C conflicts with Y: $\mathcal{T} \models D \sqcap Y \sqsubseteq \bot$. The example below illustrates the modifications that can be performed to achieve a consistent TBox.

TBox: $\{\texttt{Vehicle} \sqsubseteq \forall\texttt{Energy.Fuel} \sqcap \texttt{Moves},\ \texttt{Car} \sqsubseteq \texttt{Vehicle} \sqcap \exists\texttt{Driver.}\top\}$
New information: $\texttt{ElectroCar} \sqsubseteq \texttt{Vehicle} \sqcap \exists\texttt{Energy.}\neg\texttt{Fuel}$
Adapted TBox: $\{\texttt{Vehicle} \sqsubseteq \texttt{Moves}, \texttt{FuelVehicle} \sqsubseteq \texttt{Vehicle} \sqcap \forall\texttt{Energy.Fuel},$
$\texttt{ElectroCar} \sqsubseteq \texttt{Vehicle} \sqcap \exists\texttt{Energy.}\neg\texttt{Fuel}, \texttt{Car} \sqsubseteq \texttt{FuelVehicle} \sqcap \exists\texttt{Driver.}\top\}$

In this example, $\exists$`Energy.`$\neg$`Fuel` in the definition of `ElectroCar` conflicts with $\forall$`Energy.Fuel` subsuming `Vehicle` while `Vehicle` subsumes `ElectroCar`. In the adapted TBox the conflicting information is deleted from the definition of `Vehicle` and moved to the definition of the new concept `FuelVehicle` which is supposed to capture the original meaning of `Vehicle`. `Vehicle` needs to be replaced with `FuelVehicle` in all axioms of the TBox except for the definition of `ElectroCar`. Although we assume that the new axiom has higher priority than the original TBox, we want to keep in the definition of `Vehicle` as much information not conflicting with the definition `ElectroCar` as possible. The conflicting information is moved to the definition of the new concept `FuelVehicle`.

The presented approach works for an $\mathcal{ALE}$-TBox. Disjunction may provide a problem even on the common sense level. If an axiom $X \sqsubseteq Y$ conflicts with a disjunctive definition D of a concept in $\mathcal{T}$, then there are two possibilities to make $\mathcal{T} \cup \{X \sqsubseteq Y\}$ consistent: to add X as a new disjunct to D or to adapt one or more disjuncts in D to Y. It seems to be impossible to find a general solution for this adaptation in logics with disjunction, but the information about the conflicts may give hints to the engineer of how to modify the original TBox.

5 Adaptation Algorithm for $\mathcal{ALE}$ TBoxes

We propose a procedure adapting an $\mathcal{ALE}$-TBox $\mathcal{T}$ to a new axiom. For the sake of simplicity, assume that the set of role names contains only one name R. For a concept description C in prenex conjunctive normal form the $\mathcal{ALE}$-normal form (compare [1]) $\mathcal{ALE}$-*NF* is the conjunction of the following concepts:

- Atomic and negated atomic concepts on the top-level[6] of C (the set *prim*)
- Value restriction $\forall R.val_R(C)$, where $val_R(C) := C_1 \sqcap ... \sqcap C_n$ if $\forall i \in \{1, ..., n\} : \forall R.C_i$ occurs on the top-level of C; otherwise $val_R(C) := \top$;

[6] The concepts occurring on the top-level of the concept description C are the concepts not embedded in any relational restriction in C.

- Existential restrictions of the top-level of C;
$exr_R(C) := \{C' \mid \text{there exists } \exists R.C' \text{ on the top-level of } C\}$.

Definition 1 *The pair $\mathfrak{C} = \langle C_c, C_n \rangle$ is called conflict between the $\mathcal{ALE}$-concept descriptions C_1 and C_2, if C_c and C_n are specified as follows*

If $\exists i, j \in \{1, 2\} : C_i \equiv \top$ and $C_{j \neq i} \not\equiv \bot$ then $C_c = \{\}, C_n = \{C_1\}$;
else if $\exists i, j \in \{1, 2\} : C_i \equiv \bot$ and $C_{j \neq i} \not\equiv \bot$ then $C_c = \{C_1\}, C_n = \{\}$;
else if $C_1 \equiv C_2 \equiv \bot$ then $C_c = \{\}, C_n = \{\bot\}$; else

1. $C_c = \{C \in prim(C_1) \mid \exists C' : C' \equiv C \wedge \neg C' \in prim(C_2)\} \cup$
$\{\forall R.X \mid D \doteq val_R(C_2) \sqcap \ \bigsqcap_{D_i \in exr_R(C_2)} D_i \wedge$
$\mathfrak{C}(val_R(C_1), D) = \langle S_1, S_2 \rangle \wedge S_1 \neq \varnothing \wedge X \doteq \bigsqcap_{cc_i \in S_1} cc_i\} \cup$
$\{\exists R.X \mid \exists D \in exr_R(C_1) : \mathfrak{C}(D, val_R(C_2)) = \langle S_1, S_2 \rangle \wedge S_1 \neq \varnothing \wedge X \doteq \bigsqcap_{cc_i \in S_1} cc_i\}$
2. $C_n = \{C \in prim(C_1) \mid C \notin C_c\} \cup \{\forall R.Y \mid D \doteq val_R(C_2) \sqcap \bigsqcap_{D_i \in exr_R(C_2)} D_i \wedge$
$\mathfrak{C}(val_R(C_1), D) = \langle S_1, S_2 \rangle \ \wedge S_2 \neq \varnothing \wedge Y \doteq \bigsqcap_{cn_i \in S_2} cn_i\} \cup$
$\{\exists R.Y \mid \exists D \in exr_R(C_1) : \mathfrak{C}(D, val_R(C_2)) = \langle S_1, S_2 \rangle \wedge$
$((S_2 \neq \varnothing \wedge Y \doteq \bigsqcap_{cc_i \in S_2} cc_i) \vee Y \doteq \top))\}$

The *conflict* in Def. 1 is a tuple of two sets. The conflicting set C_c collects the concept descriptions subsuming C_1 which conflict with C_2. The non-contradicting set C_n collects all the other concept descriptions explicitly subsuming C_1.

Algorithm ***Adapt*** **for adaptation of a TBox to an axiom**
Input: `an` $\mathcal{ALE}$`-TBox` $\mathcal{T}$`, an` $\mathcal{ALE}$`-axiom` $X \sqsubseteq Y$
Output: `adapted TBox` $Adapt_{X \sqsubseteq Y}(\mathcal{T}) = \mathcal{T}'$

$\mathcal{T}' := \mathcal{T} \cup \{X \sqsubseteq Y\}$
$OC := \langle C_1, ..., C_n \rangle$: $\forall i, j \in \{1, ..., n\} : C_i \in prim(Y)$ `and` $(i < j \rightarrow \mathcal{T} \not\models C_i \sqsubseteq C_j)$
$i := 0$
`FOR` $i < n$
  $i := i + 1$; $C := C_i$: $C_i \in OC$
  `IF` $C \in ax(\mathcal{T}')$ `THEN`
    $Y \doteq C \sqcap Y', C \sqsubseteq D \in \mathcal{T}'$
    `IF` $unfolded_{\mathcal{T}'}\ D$ `and` $unfolded_{\mathcal{T}'}\ Y'$ `are` $\mathcal{ALE}$`-concept descriptions THEN`
      $\mathfrak{C}(unfolded_{\mathcal{T}'}\ D\ in\ \mathcal{ALE}\text{-}NF, unfolded_{\mathcal{T}'}\ Y'\ in\ \mathcal{ALE}\text{-}NF) = \langle C_c, C_n \rangle$
      `IF` $C_c \neq \varnothing$ `THEN`
        $\mathcal{T}' := (\mathcal{T}' \setminus (\{C \sqsubseteq D\} \cup \{Z \sqsubseteq C \sqcap Z' \in \mathcal{T}'\})) \cup$
        $\{C \sqsubseteq CN \mid CN \doteq \bigsqcap_{cu_i \in C_n} cn_i \wedge \mathcal{T}' \not\models C \sqsubseteq CN\} \cup$
        $\{C' \sqsubseteq C \sqcap CC \mid CC \doteq \ \bigsqcap_{cc_i \in C_c} cc_i\} \cup \{Z \sqsubseteq C' \sqcap Z' \mid Z \sqsubseteq C \sqcap Z' \in \mathcal{T}'\}$
`END FOR`

Algorithm *Adapt* defines an adaptation of a TBox $\mathcal{T}$ to a new axiom $X \sqsubseteq Y$ such that X becomes satisfiable towards $\mathcal{T}$. For the sake of simplicity we imply that X has not been axiomatized in $\mathcal{T}$ before, but it is easy to develop a more general procedure. For every axiomatized concept C occurring on the top-level of Y the conflict is computed for the rest of Y and the definition D of C. The conflicting concepts occurring on the top-level of D are deleted from the definition of C and moved to the definition of the new concept C' which is declared to be subsumed by C and subsume all subconcepts of C.

6 Conclusion and Future Work

We presented an approach for dynamically updating ontologies. The overall motivation is to provide a theoretical and algorithmic basis for a framework that allows to update ontological knowledge automatically handling the possible conflicts between the original ontology and the new incoming information. The main contribution of this paper is the algorithm *Adapt* adapting an $\mathcal{ALE}$-TBox to a new axiom and the notion of *conflict* (Def. 1).

We were mainly concerned with $\mathcal{ALE}$-DL, but the given approach can easily be extended to treat more expressive DLs. As sketched in Sec. 4 disjunctive definitions cannot be adapted fully automatically. Developing the prototype implementation of the *Adapt* algorithm and testing it on real data we suppose to find a solution for semi-automatic updates in the DL-logics allowing disjunction. There are two important theoretical issues we plan to examine. We will characterize the complexity of *Adapt* and the changes in the model theoretic semantics of an adapted ontology.

Acknowledgment. This line of research was partially supported by the grant MO 386/3-4 of the German Research Foundation (DFG).

References

[1] Baader, F., Küsters, R.: Non-Standard Inferences in Description Logics: The Story So Far. *International Mathematical Series*, volume 4, *Mathematical Problems from Applied Logic. New Logics for the XXIst Century* (2006) 1–75

[2] Baader, F., Lutz, C., Miličić, M. Sattler, U., Wolter, F.: Integrating Description Logics and Action Formalisms: First Results. In *Proc. of the 20th National Conference on Artificial Intelligence (AAAI'05)*, AAAI Press (2005) 572–577

[3] Baader, F., Calvanese, D., McGuinness, D., Nardi, D. Patel-Schneider (eds.), P.: *Description Logic Handbook: Theory, Implementation and Applications.* Cambridge University Press (2003)

[4] Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web – A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American (2001)

[5] Biemann, C., Shin, S., Choi, K.-S.: Semiautomatic Extension of CoreNet using a Bootstrapping Mechanism on Corpus-based Co-occurrences. *Proc. of the 20th International Conference on Comp. Ling. (Coling 2004)* (2004) 1227–1232

[6] Cohen, W., Hirsh, H.: Learning the CLASSIC Description Logic: Theoretical and Experimental Results. In *Proc. of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR94)* (1994) 121–133

[7] Fanizzi, N., Ferilli, S., Iannone, L., Palmisano, I., Semeraro, G.: Downward Refinement in the ALN Description Logic. In *Proc. of the Fourth International Conference on Hybrid Intelligent Systems (HIS'04)* (2005) 68–73

[8] Flouris, G., Plexousakis, D., Antoniou, G.: Updating Description Logics using the AGM Theory. In Proc. *The 7th International Symposium on Logical Formalizations of Commonsense Reasoning* (2005)

[9] Ghilardi, S., Lutz, C., Wolter, F.: Did I damage my ontology: A Case for Conservative Extensions of Description Logics. In *Proc. of Principles of Knowledge Representation and Reasoning 2006 (KR06)* (to appear).