Description Logic reasoning for Semantic web ontologies

– Extended abstract

Anni-Yasmin Turhan
Institute for Theoretical Computer Science
Technische Universität Dresden
Dresden, Germany
turhan@tcs.inf.tu-dresden.de

1. INTRODUCTION

The ontology language for the semantic web OWL provides means to describe entities of an application domain in an ontology in a well-structured way. The underlying formalism for OWL is Description Logics (DLs) [6], which are a family of knowledge representation formalisms that have formal semantics. This family of logics is tailored towards representing terminological knowledge of an application domain in a structured and formally well-understood way.

Description logics allow users to define important notions, such as classes or relations of their application domain in terms of concepts and roles. These concepts (unary predicates) and roles (binary predicates) then restrict the way these classes and relations are interpreted. Based on these definitions, implicitly captured knowledge can be inferred from the given descriptions of concepts and roles. These inferences are defined based on the formal semantics of DLs. A great range of these inferences has been defined and investigated for a variety of DLs with differing expressivity, where the expressivity of a DL is determined by the means it allows to describe concepts and roles. The inferences devised for DLs are, for instance, sub-class or instance relations. The investigation of algorithms for reasoning services and their complexity is the main focus of the DL research community. Mainly during the nineties reasoning procedures for more and more expressive DLs have been investigated. Naturally, the gain in expressiveness came at the cost of higher complexity for the reasoning procedures—reasoning for the DLs investigated is PSpace-complete or even ExpTime-complete [33, 28, 49] (for an overview see [9, 16]).

Despite the high complexity, highly optimized DL reasoning systems were implemented based on the tableau method—most prominently the FACT system [24] and RACER [21]. These systems employed optimization methods developed for DL reasoning based on tableaux [7, 23, 32, 22] and demonstrated that the high worst case complexities would hardly be encountered in practice [24, 32, 20, 25]. In fact, it turned out that these highly optimized implementations of the reasoning methods do perform surprisingly well on DL knowledge bases from practical applications. Encouraged by these findings and driven by application needs researchers investigated tableau algorithms for even more expressive DLs [29, 30, 26, 31] in the last decade, which were implemented in tableau-based reasoners for very expressive DL such as PELLET [45], FACT++ [50, 51] and RACERPRO [44]. More recently, new reasoning methods for expressive DLs were investigated and implemented such as resolution [37, 39] in KAON2 and hyper-tableau [40, 41, 19] in HERMIT.

Another line of research is dedicated to the design of so-called light-weight DLs, which are DLs with relatively limited expressivity, but good computational properties for specific reasoning tasks [8]. Reasoning even for very large ontologies written in these DLs can be done efficiently, since the respective reasoning methods are tractable. There are two families of light-weight DLs: the $\mathcal{EL}$ family [12, 4, 5], for which the subsumption and the instance problem are polynomial, and the DL Lite family [13, 15], for which the instance problem and query answering are polynomial. A member of each of these families is the DL corresponding to one of the profiles of the OWL 2 standard.

DLs have been employed in various domains, such as databases, biomedical or context-aware applications [3, 48]. Their most notable success so far is probably the adoption of the DL-based language OWL as standard ontology language for the Semantic Web [27].

In the following we give pointers to the relevant literature for the topics that were covered in the tutorial.

2. DL PRELIMINARIES

Historically, DLs stem from knowledge representation systems such as semantic networks [43, 46] or frame systems [36]. These early knowledge representation systems were motivated by linguistic applications and offered methods to compute inheritance relations between the specified notions from the domain of discourse. These early knowledge representation systems have operational semantics, i.e., the semantics of reasoning is given by its implementation. As a consequence, the result of the reasoning process depends on the implementation of the reasoner [47]. To remedy this, DLs and their reasoning services are based on formal semantics. The information about the application domain is represented in a declarative and unambiguous way. More importantly, the formal semantics of the reasoning services ensure behavior of the DL reasoning systems independent of their implementation. For comprehensive introductions on DLs we refer the reader to [6, 2, 52].

The central notion for DLs are concept descriptions, which can be derived from concept names and so-called concept constructors. For instance, one can describe a tutorial as an event given by a lecturer in the following way by a concept description:

\[
\text{Event} \sqsubseteq \text{given-by.Lecturer} \sqcap \exists \text{has-topic.T}
\]
This concept description is a conjunction ($\land$) of the concept Event, the existential restriction $\exists$ has-topic $\top$, and the existential restriction $\exists$ has-topic $\top$. The first existential restriction consists of the role name given-by and concept Lecturer, which relates the Lecturer to the tutorial. The latter existential restriction states that there is a topic (which is not specified).

In general, concept descriptions are built from the set of concept names $N_C$ and the set of role names $N_R$ using concept constructors. Every DL offers a different set of concept constructors. The DL $\mathcal{EL}$ allows only for the concept constructors that were used in the example concept description above. We call concept descriptions of the form $\exists r.C$ existential restrictions and concept descriptions of the form $C \land D$ conjunctions. The semantics of DL concept descriptions is given by means of interpretations.

Let $C$ and $D$ be $\mathcal{EL}$-concept descriptions and $r$ a role name. An interpretation is a pair $I = (\Delta^I, \mathcal{I})$ where the domain $\Delta^I$ is a non-empty set and $\mathcal{I}$ is a function that assigns to every concept name $A$ a set $A^I \subseteq \Delta^I$ and to every role name $r$ a binary relation $r^I \subseteq \Delta^I \times \Delta^I$. This function is extended to complex $\mathcal{EL}$-concept descriptions as follows:

- $(C \land D)^I = C^I \land D^I$;
- $(\exists r.C)^I = \{ x \in \Delta^I \mid \text{there is } y \in \Delta^I \text{ with } (x, y) \in r^I \}$ and $y \in C^I$.

Most DLs also offer the top-concept $\top$, which is always interpreted as the whole domain $\Delta^I$. Now, with the $\mathcal{EL}$-concept constructors at hand, one can, for instance, characterize a graduate CS student by the following concept description:

$\exists$ studies. CS $\land$ (Master-Student $\land$ has-degree. Bachelor)

Concept description like these are the main building blocks to model terminological knowledge. The DL defined above is called $\mathcal{EL}$ and is the core DL of the $\mathcal{EL}$-family.

Depending on the set of concept constructors on can define more expressive DLs. The concept constructors given in Table 1 form the DL $\mathcal{ALCQ}$.

For expressive DLs, roles can have declared properties or can be related to other roles. A role $r$ can be declared to be a transitive role in the $\mathcal{TBox}$. The semantics is straightforward. An interpretation $I$ satisfies a transitive role declaration transitive($r$) if $(a, b), (b, c) \subseteq r^I$ implies $(a, c) \in r^I$. The declaration of an inverse role applies to a role name $r$ and yields its inverse $r^{-1}$, where the semantics is the obvious one, i.e., $(r^{-1})^I := \{(e, d) \mid (d, e) \in r^I\}$. Furthermore, it can be specified that a role is a superrole of another role by a role inclusion axiom. The set of all role inclusions forms the role hierarchy. An interpretation $I$ satisfies a role inclusion axiom $r \sqsubseteq s$ if $r^I \subseteq s^I$. The DL that allows all the concept constructors and role declarations introduced is the DL $\mathcal{SHIQ}$. This DL is the logic underlying OWL DL.

DL knowledge bases are typically divided two parts. The

### Terminological part

which captures the characterization of the basic categories from an application, and the

### Assertional part

which captures the facts from the application.

We introduce these components of DL knowledge bases now.

#### 2.1 Terminological Knowledge

A name can be assigned to a concept description by a concept definition. For instance, we can write $\text{Tutorial} = \text{Event} \land \exists \text{given-by} \ \text{Lecturer}$ to supply a concept definition for the concept Tutorial. Let $A$ be a concept name and $C$, $D$ be (possibly) complex concept description.

- A concept definition is a statement of the form $A \equiv C$.
- A general concept inclusion (GCI for short) is a statement of the form $C \sqsubseteq D$.

It is easy to see that every concept definition $A \equiv C$ can be expressed by two GCIs: $A \sqsubseteq C$ and $C \sqsubseteq A$. The terminological information expressed by GCIs is collected in the so-called $\mathcal{TBox}$, which is simply a finite set of GCIs. An interpretation is a model of a $\mathcal{TBox}$, if it satisfies all GCIs, i.e., if $C^I \subseteq D^I$ for all $C, D \in \mathcal{T}$.

If all concept descriptions in a $\mathcal{TBox}$ are from a description logic $\mathcal{L}$, then we call $\mathcal{T}$ a $\mathcal{L}$-TBox.

If a concept description $A \equiv C$ in a $\mathcal{TBox}$ uses a concept name $B$ directly, i.e., $B$ appears in $C$, or if $B$ is used indirectly by the definitions of the names appearing in $C$, we say that the $\mathcal{TBox}$ is cyclic. Otherwise a $\mathcal{TBox}$ is acyclic.

#### 2.2 Assertional Knowledge

Individual facts from the application domain can be stated by assertions. There are two basic kinds of assertions for DL systems—one expresses that an individual belongs to a concept and the other one specifies that two individuals are related via a role. The set $N_I$ is the set of all individual names. Let $C$ be a (possibly complex) concept description, $r \in N_R$ a role name and $i, j \in N_I$ be two individual names, then

- $C(i)$ is called a concept assertion and
- $r(i, j)$ is called a role assertion.

An $\mathcal{ABox} \mathcal{A}$ is a finite set of concept assertions and role assertions.

For instance, we can express that Dresden is a city located at the river Elbe by the following $\mathcal{ABox}$:

{ City(Dresden), River(Elbe), located-at(Dresden, Elbe) }

If all concept descriptions in an $\mathcal{ABox} \mathcal{A}$ are from a Description Logic $\mathcal{L}$, then we call $\mathcal{A}$ a $\mathcal{L}$-$\mathcal{ABox}$.

In order to capture $\mathcal{ABoxes}$, the interpretation function is now extended to individual names. Each individual name is mapped by the interpretation function to an element of the domain $\Delta^I$. Let $C$ be a concept description, $r$ a role name and $i, j$ two individual names, then an interpretation $I$ satisfies

### Table 1: OWL syntax, DL syntax and semantics of concept descriptions.

<table>
<thead>
<tr>
<th>constructor name</th>
<th>OWL syntax</th>
<th>DL syntax</th>
<th>semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>conjunction</td>
<td>intersectionOf</td>
<td>$C \land D$</td>
<td>$C^I \land D^I$</td>
</tr>
<tr>
<td>existential restriction</td>
<td>someValuesFrom</td>
<td>$\exists r.C$</td>
<td>${ x \in \Delta^I \mid # { y \mid (x, y) \in r^I \land y \in C^I } }$</td>
</tr>
<tr>
<td>value restriction</td>
<td>allValuesFrom</td>
<td>$\forall r.C$</td>
<td>${ x \in \Delta^I \mid \forall y : (x, y) \in r^I \rightarrow y \in C^I }$</td>
</tr>
<tr>
<td>negation</td>
<td>complementOf</td>
<td>$\neg C$</td>
<td>$\Delta^I \setminus C^I$</td>
</tr>
<tr>
<td>disjunction</td>
<td>unionOf</td>
<td>$C \lor D$</td>
<td>$C^I \lor D^I$</td>
</tr>
<tr>
<td>qualified number restrictions</td>
<td>qualified cardinality restrictions</td>
<td>$(\leq \ n \ r C) \</td>
<td>{ x \in \Delta^I \mid # { y \mid (x, y) \in r^I \land y \in C^I } \leq n }$</td>
</tr>
</tbody>
</table>
• the concept assertion $C(i)$ if $i^2 \in C^2$ and

• the role assertion $r(i, j)$ if $(i^2, j^2) \in r^2$.

An interpretation $I$ is a model of an ABox $A$, if $I$ satisfies every assertion in $A$. A DL knowledge base $K$ consists of an ABox $A$ and a TBox $T$. We write $K = (T, A)$. We now discuss standard reasoning services for DLs, which are implemented in most DL systems.

### 2.3 Reasoning services

One of the basic reasoning services in DL systems is to test for the satisfiability of a concept or a TBox, i.e., to test whether the information specified in it contains logical contradictions or not. In case the TBox contains a contradiction, any consequence can follow logically from the TBox. Moreover, if a TBox is not satisfiable, the specified information can hardly capture the intended meaning from an application domain. To test for satisfiability is often the first step for a user to check whether a TBox models something “meaningful”.

Let $C$ be a concept description and $T$ a TBox. The concept description $C$ is satisfiable iff it has a model, i.e., iff there exists an interpretation $I$ such that $C^I \neq \emptyset$. A TBox $T$ is satisfiable iff it has a model, i.e., an interpretation that satisfies all GCI’s in $T$.

If a concept or a TBox is not satisfiable, it is called unsatisfiable. Other typical reasoning services offered in DL systems test for equivalence or inclusion relations between concepts. In the latter case, if one concept of the TBox models a more general category than another one, we say that this concept subsumes the other one.

Let $C, D$ be two concept descriptions and $T$ a (possibly empty) TBox. The concept description $C$ is subsumed by the concept description $D$ w.r.t. $T$ ($C \sqsubseteq_D T$), iff $C^2 \subseteq D^2$ holds in every model $I$ of $T$. Two concepts, $C, D$ are equivalent w.r.t. $T$ ($C \equiv_T D$), iff $C^2 = D^2$ holds for every model $I$ of $T$.

A test for the equivalence of concept descriptions can be carried out by two subsumption tests, since the following holds: ($C \equiv_T D$), iff $C \sqsubseteq_T D$ and $D \sqsubseteq_T C$.

The computation of the subsumption relations for all named concepts mentioned in the TBox $T$ is called classification of the TBox $T$ and yields the concept hierarchy of the TBox $T$.

Similarly as for concept descriptions and TBoxes, we can test for the absence of contradictions in ABoxes. An ABox $A$ is consistent w.r.t. a TBox $T$, iff it has a model that is also a model for $T$. The individual $i$ is an instance of the concept description $C$ w.r.t. an ABox $A$ and a TBox $T$ (we write $A \models_T C(i)$), iff $i^2 \in C^2$ for all models $I$ of $T$ and $A$. The reasoning service ABox realization computes for each individual $i$ of an ABox $A$ and a TBox $T$ the set of all named concepts $A$ appearing in $A$ and $T$ that (1) have $i$ as an instance ($A \models_T A(i)$) and (2) that is least w.r.t. $\sqsubseteq_T$.

With the help of the instance service, one can also compute answers to instance queries, i.e., all individuals occurring in the ABox that are instances of the query concept $C$. In order to state more general search criteria, one can use so-called conjunctive queries, i.e., conjunctions of assertions that may also contain variables, of which some can be existentially quantified. For example, the conjunctive query

\[ \exists x, y. \text{City}(x) \land \text{River}(y) \land \text{located-at}(x, y) \]

asks for all Cities that located at a river. With respect to the ABox we have introduced earlier, this conjunctive query has the pair of individuals: Dresden, Elbe as an answer.

### 2.4 Reasoning Techniques

There are three main reasoning approaches for the DLs that underlie OWL. For the expressive DLs, which offer all Boolean concept constructors, most reasoning services can be reduced to consistency of an ABox w.r.t. a TBox in polynomial time. In presence of full negation we can devise the following polynomial time reductions:

• Equivalence can be be reduced to subsumption: $C \equiv_T D$ iff $C \sqsubseteq_T D$ and $D \sqsubseteq_T C$.

• Subsumption can be be reduced to (un)satisfiability: $C \sqsubseteq_T D$ iff $C \cap \neg D$ is unsatisfiable w.r.t. $T$.

• Satisfiability can be be reduced to consistency: $C$ is satisfiable w.r.t. $T$ iff the ABox $\{C(a)\}$ is consistent w.r.t. $T$.

• The instance problem can be reduced to (in)consistency: $A \models_T C(a)$ if $A \cup \{\neg C(a)\}$ is inconsistent w.r.t. $T$.

As a consequence, only one reasoning procedure needs to be investigated and implemented in systems in order to use all reasoning services in practice. In fact, all reasoning systems for expressive DLs use this approach and implement the tableau method for the consistency test. For a more detailed discussion see [2, 52]. Even query answering can be implemented by the tableau method [18, 34, 42, 17]. However, the reduction-based approach requires the presence of negation in the DL.

For light-weight DLs this constructor is not available. Thus different methods for reasoning have been devised. In case of the EL-family, the completion method [12, 4, 5] actually classifies the given TBox $T$, i.e., it simultaneously computes all subsumption relations between the concept names occurring in $T$. This algorithm proceeds in four steps:

1. Normalize the TBox.
2. Translate the normalized TBox into a graph.
3. Complete the graph using completion rules.
4. Read off the subsumption relationships from the normalized graph.

By the completion method a TBox can be classified in polynomial time. This method is implemented in the EL-reasoner jCEL [35].

The other lightweight family of DLs is the DL-Lite family, which is tailored to perform query answering efficiently [14, 15]. Here the approach is to translate the initial query such that it also captures the relevant information from the TBox and then use a relational query engine to answer the translated query. For a comprehensive description of the method see [8]. This method for query answering is implemented in the system QuOnto [1].

### 3. THE STANDARD ONTOLOGY LANGUAGE OWL

The first version of the OWL standard [27, 10] covered three dialects of increasing expressivity. While the most expressive dialect OWL full is beyond the expressivity of DLs and reasoning in it is undecidable, the other two dialects correspond to DLs for which sound and complete reasoning procedures exist. OWL DL can express ontologies written in the DL SHOIQ and the less expressive OWL lite can express ontologies written in the DL SHWe.

The close relation of DLs and these two OWL dialects raised the interest of new user groups in DL reasoning and DL systems on the
REFERENCES


1998.


