

Description Logics for the Semantic Web: Racer as a Basis for Building Agent Systems

Ralf Möller, Univ. of Applied Sciences of Wedel, 22880 Wedel, Feldstr. 143
Volker Haarslev, Concordia University, Comp. Sc. Dept., 1455 de Maisonneuve Blvd. W.,
Montreal, Quebec H3G 1M8, Canada

The term Semantic Web denotes a vision of a new World-Wide Web in which different kinds of resources (data, services, Web pages, etc.) are accessed and shared on the basis of formal representation structures [6]. As a well-motivated design decision and in order to integrate Semantic Web facilities into the existing software infrastructure of the current Web, its main XML-based syntactic representation formats and their processing conventions are taken for granted. However, in contrast to the standard way of accessing current Web resources by means of surface-oriented syntactic comparison operations, the main idea of the Semantic Web is to answer different kinds of queries w.r.t. a specific conceptual data model rich enough in expressivity to adequately describe the interrelationships between basic representational units (names, symbols) on which a piece of software for a certain application domain is based. Since symbols might have different meanings in different software applications, it is important to explicitly refer to meaning definitions given as a certain instance of an underlying conceptual data model. In the context of the Semantic Web, these definitions are seen as resources themselves and, for reasons of brevity, are often called ontologies. Ontologies provide the basis for implementing specific Semantic Web resources for encapsulating computational processes. These resources are called services [18]. Thus, ontologies play an important role for the Semantic Web. But who is responsible for building ontologies? Influenced by experiences with previous artificial intelligence approaches and motivated by the success of the current Web, many researchers argued for a decentralized logic-based approach [6], and it is the idea of a formal semantics on which the logic is based that led to the name of the proposed new medium: Semantic Web.

A logical semantics provides the basis not only for the meaning of the representational language but also for the meaning of the associated query and command language. In many applications the logic behind conceptual data models must be decidable but very expressive [1] such that, in principle, query answering involves solving nontrivial logical inference problems. This article briefly introduces the foundation of Semantic Web ontology representation languages, namely description logics [2]. For actually building systems for the Semantic Web, however, we consider practical aspects of very expressive inference systems and their integration into a distributed systems context. Using the Racer System [15] as an example the paper sketches how description logic inference engines can be used to implement information retrieval services on the Semantic Web. The article is written for Semantic Web system developers who would like to learn about logic-based software infrastructure available off the shelf and ready for implementing ontology-based services. We will demonstrate the power of description logics and explain why they are particularly suited for the Semantic Web. However, details on the semantics of description logics in general (and the one underlying Racer in particular), are not to be repeated here (see [1, 16, 11]). We assume the reader is largely familiar with XML Web standards and conceptual modeling techniques well-known in the database field.

Agents, Ontologies, and Problem-Solving

The demand for more and more flexible software systems led to the emergence of the agent paradigm for software development [21]. In particular, in the context of the Semantic Web,

agent systems offering and utilizing services are an integral part of the grand vision. The underlying principles for agents are rationality and autonomy [21, p. 31ff.]. The development of autonomous agents for interacting with software systems in foreign environments is a difficult problem, and there are three different perspectives on agent systems that are to be discussed here.

First of all, almost all software systems, including agent systems, are based on some assumptions about the meaning of names. Hence, for constructing an agent system, the names used in the agent architecture itself must be systematically related to one another in order to ensure the development of a system architecture that is understandable, extensible, adaptable, etc. This is known as conceptual data modeling (with conceptual notions such as generalization, disjointness, coverings for classes and, for instance, generalization as well as domain and range restrictions for relations). Conceptual data modeling is important at agent development time [7].

Second, and probably even more challenging, there is the problem of relating the names of the terminology internally used by an agent to the names used in the host environment the agent is eventually interacting with. If an agent is supposed to successfully interact with some foreign software system providing a host environment not anticipated at agent construction time, then the problem of assigning meaning to names is prevalent since the meaning is subject to explicit computational processing at agent runtime (see [8]).

Third, the ultimate goal could be to implement (parts of) problem-solving processes of agents such as, for instance, information retrieval, planning, scheduling, controlling, negotiation etc. on the basis of reasoning processes, i.e. subproblems of applications are to be modeled as logical decision problems. Based on a sound logical basis, correctness of (parts of) problem-solving processes can be guaranteed, and more powerful systems can be built with less resources. Needless to say that specific ontologies are required for representing problem-solving knowledge and for supporting adequate logical formalizations of application problems to be solved when the agent is in operation [2].

In all three views discussed above, ontologies are a prerequisite for agent problem solving on the Semantic Web. In the next section we give an overview about the logical background for defining ontologies.

Description Logics for Specifying Ontologies and Knowledge Bases

The first decision to be made for developing an ontology concerns so-called ontological commitments [21, p. 165]. What is the assumed nature of reality? Experiences from decades of research and practice indicate that a view in which the world consists of objects and relations between them provides an appropriate basis for many applications. Description logics exactly support this view. Concept names denote sets of objects from a so-called universe of discourse, role names denote sets of tuples of objects. For historical reasons, (binary) relations between objects are called roles, the objects on the righthand side of a tuple in a relation are called role fillers. In order to demonstrate the benefit of description logics for developing ontologies we consider a very basic scenario everyone should be able to intuitively understand.

First of all, for a particular ontology the names for basic concepts and roles have to be chosen. In our example, we consider concept names such as HUMAN, PARENT, WOMAN, MAN,

AUDI, BWM, MERCEDES, TAXI, and so on. In addition, role names used in our example ontology are the following: *has-child*, *has-part*, *owns*, etc.

In order to adequately describe sets of objects for certain application purposes (see below for an agent information retrieval application), names can be combined using concept construction operators to form new concepts. For instance, the concept “a man who has a child which is a parent” can be specified as

$$\text{MAN} \sqcap (\exists \textit{has-child} \text{ PARENT})$$

Note that for all objects which are member of the set of objects described by this concept, the child has to exist (and it must be a person). Sometimes it is not the goal to specify the existence of a filler for a certain role but rather to specify restrictions for all fillers of a role (if there are any). For instance, “women all of whose children are men” can be described with

$$\text{WOMAN} \sqcap (\forall \textit{has-child} \text{ MAN})$$

Note that the restriction is met if the woman has no children at all. Another example demonstrates the use of negation. The catholic church still forbids female popes such that $\neg \text{WOMAN}$, the complement of the set denoted by WOMAN, might be part of a concept for describing popes. Concepts can be recursively combined. For instance, the notion of “a human who owns an AUDI, a BMW or a MERCEDES” can be specified using the expression

$$\text{HUMAN} \sqcap (\exists \textit{owns} (\text{AUDI} \sqcup \text{BMW} \sqcup \text{MERCEDES}))$$

The nested concept ($\text{AUDI} \sqcup \text{BMW} \sqcup \text{MERCEDES}$) also demonstrates the representation of disjunctive or underspecified information. Another example could be a “woman with at least 2 female children but not more than 4 children overall”. This can be described with the concept

$$\text{WOMAN} \sqcap (\geq 2 \textit{has-child} \text{ WOMAN}) \sqcap (\leq 4 \textit{has-child})$$

So far we have used concepts for *describing* sets of objects. However, for specifying ontologies it is also important to relate different sets of objects to each other, for example w.r.t. the subset relationship. This can be done using so-called terminological axioms. The simplest example uses just names:

$$\text{MAN} \sqsubseteq \text{HUMAN} \quad \text{and} \quad \text{WOMAN} \sqsubseteq \text{HUMAN}$$

However, ontological modeling in description logics is not limited to using names only. We might want to enforce that all taxis are autos with at least 4 seats.

$$\text{TAXI} \sqsubseteq \text{AUTO} \sqcap (\geq 4 \textit{has-part} \text{ SEAT})$$

$$\text{SMART} \sqsubseteq \text{AUTO} \sqcap (\leq 2 \textit{has-part} \text{ SEAT})$$

The axiom specifies necessary restrictions for all taxis and Smarts. Once it is known that an object is a taxi, it becomes clear that it has at least 4 seats. On the other hand, an auto with at least 4 seats is not necessarily a taxi. Similar restrictions are imposed for Smarts. As another example a necessary restriction for all parents is imposed as follows:

$$\text{PARENT} \sqsubseteq \text{HUMAN} \sqcap (\exists \textit{has-child} \text{ HUMAN})$$

Disjointness of concepts can easily be declared with axioms of the form:

$$\text{MAN} \sqsubseteq \neg \text{WOMAN}$$

The set of men is included in the set of non-women and, therefore, it is disjoint from the set of women. Coverings can also be expressed.

$$\text{HUMAN} \sqsubseteq (\text{WOMAN} \sqcup \text{MAN})$$

The set of humans is covered by the union of the sets with names WOMAN and MAN. Other conceptual modeling tasks can also be accomplished using terminological axioms. For instance, once we know that a man has a child which is a parent, this is also sufficient to call this man a grandfather.

$$\text{HUMAN} \sqcap (\exists \textit{has-child} \text{ HUMAN}) \sqsubseteq \text{PARENT}$$

The axiom can be paired with the above axiom for PARENT. Since such a pair of axioms

$C \sqsubseteq D$ and $D \sqsubseteq C$ is quite common in ontology modeling, the shorthand $C \equiv D$ for both axioms is often used as follows.

$$\text{PARENT} \equiv \text{HUMAN} \sqcap (\exists \textit{has-child} \text{ HUMAN})$$

In a similar way necessary and sufficient conditions for the concept UNCLE can be specified:

$$\text{UNCLE} \equiv \text{MAN} \sqcap (\exists \textit{has-sibling} \text{ PARENT})$$

Using description logics, terminological knowledge for ontologies can be specified in an adequate way for many applications. Note that it is absolutely possible to specify sufficient conditions only. For instance, if it is known that for someone there exists a child which is a human, then this someone must be a mother or a father.

$$(\exists \textit{has-child} \text{ HUMAN}) \sqsubseteq (\text{MAN} \sqcup \text{WOMAN})$$

In ontologies, not only concepts are important, restrictions on roles have to be specified as well. For instance, one might want a role (relation) *has-mother* to be functional because only one mother should be assigned to each object. The role *has-descendant* might be defined as transitive, and the role *has-child* should probably be a subrole of *has-descendant* because every tuple in the relation *has-child* should also be in the relation *has-descendant*. The relation *has-descendant* is the inverse of *has-ancestor*. In addition, for our ontology we might assume that for the relation *has-child*, the domain and range are PARENT and HUMAN respectively (see also [3] for many examples involving the above-mentioned modeling constructs).

In description logics, for historical reasons, a set of axioms of these kinds is also referred to as a T-box. Using T-boxes adequate conceptual notions required for ontologies can be defined. In terms of description logics, the name ontology is often used as a synonym for T-box.

Dealing with sets of objects and T-boxes for conceptual modeling we have painted only half of the complete picture of description logics. For agent problem solving, referring to single objects and tuples of single objects is equally important, of course. Names for single objects are to be specified as well (these names are usually called individuals). In our family domain,

we might need individuals such as ALICE, BETTY, etc. For declaring knowledge about individuals so-called assertional axioms are provided. If we assume that ALICE is a woman, this can be stated as

ALICE : WOMAN

With an assertion of the form

(ALICE, BETTY) : *has-child*

individuals of our family domain can be related to one another.

A set of these assertions is called an A-box. A pair (T, A) of a T-box T and an A-box A is called a knowledge base. Unfortunately, in the literature, the name ontology is often also used to refer to a knowledge base. The expressiveness of the description logic discussed in this section is usually sufficient for many practical applications. In addition to the operators sketched above, many other representations techniques have been investigated in the field of description logics as well. For instance, using algebraic constructs one can easily define people with fever as humans who have a temperature of over 40 degrees Celsius. Since in the context of the Semantic Web, units might vary, having the linear relationship between Celsius and Fahrenheit correctly specified might allow agents from different countries to effectively interact. For the sake of completeness, it should be mentioned that the description logic we refer to is usually called SHIQ(\mathcal{D}_n) (cf., [16] and [11]), but due to space restrictions we cannot explain all features in detail (see also [3] for many additional examples).

Reasoning and Inference Problems

Given a T-box, various kinds of queries can be answered. Based on the logical semantics of the representation language, different kinds of queries are defined as inference problems (hence, answering a query is called providing inference service). As a summary, we list only the most important ones here and discuss examples in the context of an agent scenario later on:

- Concept consistency w.r.t. a T-box: Is the set of objects described by a concept empty?
- Concept subsumption w.r.t. a T-box: Is there a subset relationship between the set of objects described by two concepts?
- Find all inconsistent concepts mentioned in a T-box. Inconsistent concepts might be the result of modeling errors.
- Determine the parents and children of a concept w.r.t. a T-box: The parents of a concept are the most specific concept names mentioned in a T-box which subsume the concept. The children of a concept are the most general concept names mentioned in a T-box that the concept subsumes. Considering all concept names in a T-box the parent (or children) relation defines a graph structure which is often referred to as taxonomy. Note that some authors use the name taxonomy as a synonym for ontology.

Note that whenever a concept is needed as an argument for a query, not only predefined names are possible. If also an A-box is given, among others, the following types of queries are possible:

- Check the consistency of an A-box w.r.t. a T-box: Are the restrictions given in an A-box w.r.t. a T-box too strong, i.e., do they contradict each other? Other queries are only possible w.r.t. consistent A-boxes.
- Instance testing w.r.t. an A-box and a T-box: Is the object for which an individual stands a member of the set of objects described by a certain query concept? The individual is then called an instance of the query concept.
- Instance retrieval w.r.t. an A-box and a T-box: Find all individuals from an A-box such that the objects they stand for can be proven to be a member of a set of objects described by a certain query concept.
- Computation of the direct types of an individual w.r.t. an A-box and a T-box: Find the most specific concept names from a T-box of which a given individual is an instance.

With the inference services introduced above, conceptual domain models can be evaluated and application problems can be solved. Description logics provide a logical language which is decidable, i.e., algorithms for solving inference problems exist. In particular, description logics stand for fragments of first-order logic that exhibit quasi-tractable behavior. Despite the exponential worst-case complexity for very expressive variants of description logics, practical experiences with the average-case behavior of highly optimized description logic systems indicate that many practical problems can be solved with adequate response times. Nevertheless, only highly optimized and well-tested systems such as Racer can cope with realistic T-boxes and A-boxes. In addition, to be practically useful, description logics have to be integrated into current Web software architectures. We first turn to syntax issues and deal with distributed access to inference services afterwards.

Description Logics and Semantic Web Representation Languages

One of the standards for the Semantic Web is the Resource Description Format (RDF, [17]). Since RDF is based on XML it shares its document-oriented view of grouping sets of declarations or statements. With RDF's triple-oriented style of data modeling, it provides means for expressing graph-structured data over multiple documents (whereas XML can only express graph structures within a specific document). As a design decision, RDF can talk about everything. Hence, in principle, statements in documents can also be referred to as resources. In particular, conceptual domain models can be represented as RDF resources. Conceptual domain models are referred to as "vocabularies" in RDF. Specific languages are provided for defining vocabularies (or ontologies). An extension of RDF for defining ontologies is RDF Schema (RDFS [20]) which only can express conceptual modeling notions such as generalization between concepts (aka classes) and roles (aka properties). For properties, domain and range restrictions can be specified. Thus, the expressiveness of RDFS is very limited. Much more expressive representation languages are DAML+OIL [10] and OWL [19]. Although still in a very weak way, based on XML-Schema, DAML+OIL provides for means of dealing with algebraic data types such as numbers.

DAML+OIL also allows for individuals in concepts (and T-box axioms). For example, expressing the fact that all human stem from a single human called ADAM requires an individual in a concept (and a T-box):

$$\text{HUMAN} \sqsubseteq (\exists \textit{has-ancestor} \{\text{ADAM}\}).$$

Neglecting individuals in concepts, the logical basis of DAML+OIL could be characterized with the logic SHIQ(Dn)-, i.e., with some restrictions, DAML+OIL documents can be automatically translated to SHIQ(Dn)- T-boxes [3]. The RDF-Part of DAML+OIL documents

which does not refer to schema knowledge can be translated to SHIQ(Dn)- A-boxes (this can be seen as an abstraction w.r.t. meta-level reasoning but we cannot go into details here).

Building and Processing Resources Specified in RDF or DAML+OIL

As we have discussed in the introduction, for really building systems for the Semantic Web, it is also important to consider practical description logic inference systems. To be more concrete, we concentrate on the first system supporting A-boxes and very expressive description logics, namely Racer [31, 20], and discuss its capabilities for the Semantic Web. Racer is available as an application program providing a Semantic Web inference server (called the Racer Server). Based on TCP and HTTP, Racer is integrated into the distributed systems context on various platforms (Linux, Mac OS X, Windows).

One proposal for a query language for RDF resources and for interconnecting inference engines is specified in the HTTP-based DIG standard (see [5]). DIG provides an XML-based language for defining terminological knowledge, for defining knowledge about individuals, for specifying queries, and for receiving answers to queries. With its socket-based TCP interfaces, Racer extends the DIG standard by supporting many additional query and declaration possibilities [12]. Following the DIG protocol or Racer's native TCP protocol, agent programs can submit RDF/DAML+OIL documents to the Racer Server and pose queries. APIs for Java and C++ are available.

In order to make description logics practical, graphical interfaces are of utmost importance for building ontologies. One powerful system that can be used for graphically building ontologies and maintaining DAML+OIL documents is OilEd [4]. OilEd is developed by Sean Bechhofer from Manchester University. Obeying the DIG protocol and using Racer (among others), OilEd can verify the consistency of concepts, find implicit generalization relationships in the domain model, determine the direct type(s) of individuals, etc.

Let us assume that the T-box introduced above is specified with OilEd. We assume that necessary and sufficient condition for a concept BROTHER is specified as follows by an OilEd user:

$$\text{BROTHER} \equiv \text{MAN} \sqcap (\exists \textit{has-sibling} \text{HUMAN})$$

OilEd can employ the description logic system RACER to automatically detect that the concept BROTHER is subsumed by the concept UNLCE already defined in the knowledge base. Although syntactically not directly apparent in the axiom above, the reason is that parents are also humans (compare the axiom for UNCLE presented above).

Furthermore, if an ontology designer specifies

$$\text{ECONOMIC_TAXI} \equiv \text{TAXI} \sqcap \text{SMART}$$

RACER reveals an inconsistency. Due to the example T-box considered here, there can be no economic taxis as specified by this axiom because Smarts have at most 2 seats whereas taxis require at least 4 (see above). Inconsistent concept names are usually due to modeling errors. Why should one give a (more or less) complex axiom for something inconsistent? Hence, it becomes clear that description logics in general, and Racer in particular, can be used for systematically building domain models.

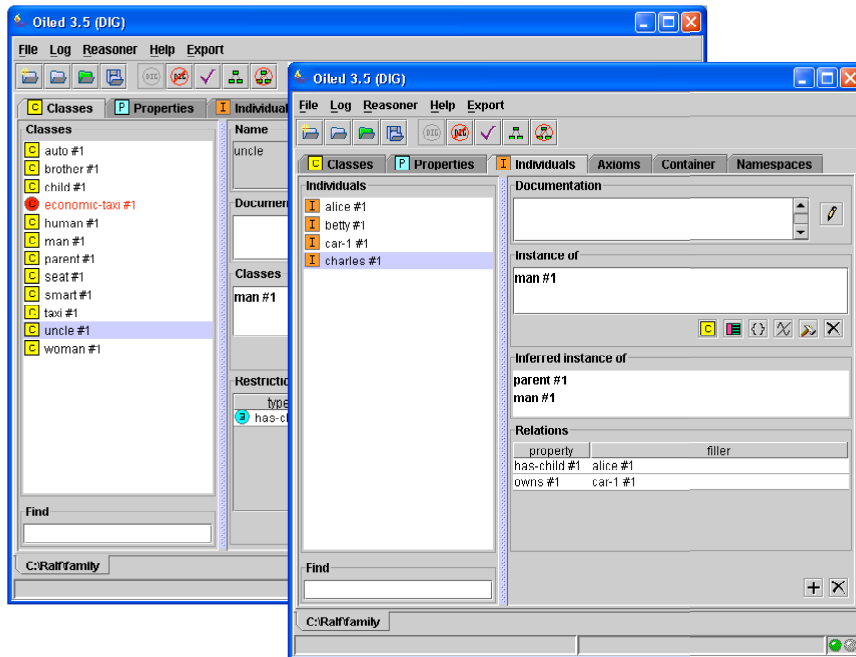


Figure 1: An example ontology defined with OilEd.

Another flexible graphical interface for managing ontologies with Racer is RICE (Racer Interactive Client Environment [9]). RICE is developed by Ronald Cornet, University of Amsterdam, Department of Medical Informatics. Based on the TCP interface of Racer, RICE provides full access to all inference services provided by Racer (see Figure 2).

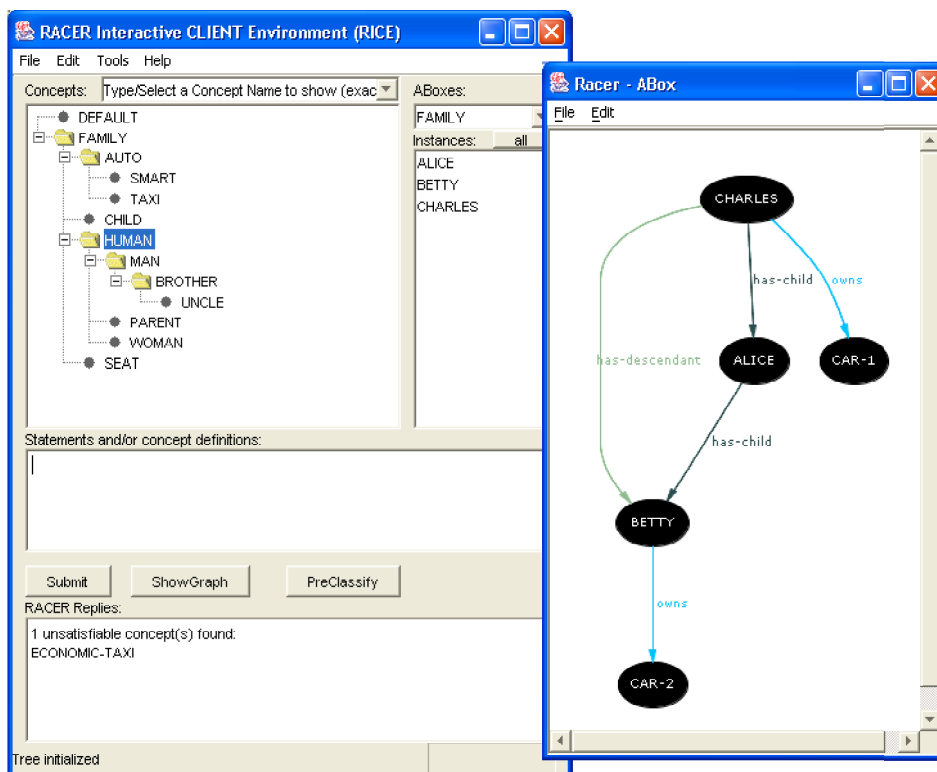


Figure 2: Snapshot of the RICE interface.

It should be clear that the use of description logics is not limited to system development time. At runtime, applications can also directly connect to a Racer Server and solve application problems using logical queries. This is discussed in the next section.

A Vision for Building Agent Systems with a Description Logic Inference Engine

In the following we consider an information retrieval scenario in which we assume that agents migrate to a specific agent host environment [21, p. 45] in order to retrieve information for a certain client. Based on certain application-specific epistemological decisions [21, p. 165] agents and host environments can be built using description logic inference services. We assume that information to be retrieved is stored in the host environment in the form of a description logic A-box. The background knowledge (conceptual data model) for retrieval tasks is given in the form of description logic T-boxes. Motivated by the idea of the Semantic Web, the T-boxes and A-boxes involved are assumed to stem from resources specified in DAML+OIL and RDF documents, which are stored in the agent host environment (or are retrieved on demand). If we assume that the agent host environment provides a Racer Server operating over HTTP or TCP, the implementation of the scenario with a Java-based software infrastructure (e.g., Voyager for object migration) becomes more than realistic. The information retrieval service is assumed to be registered in an appropriate way [18] such that agents can find the service, can migrate to the agent host environment on the information server and pose queries.

Within the above-mentioned framework, different scenarios are possible: First, the host environment can be tailored to a specific class of agents, i.e., the agents and the host environment share the same ontology (T-box). Second, and probably more challenging, information retrieval agents and host environments use different ontologies (T-boxes) and employ a third instance, a so-called ontology agent, to provide so-called inter-schema translation facilities.

Agents and Host Environment Share the Same Ontology

We assume that the T-box of the agent and the host environment consists of the axioms discussed above. In addition to the A-box presented above, we assume that the A-box (of the host environment) contains some additional assertions:

CHARLES : MAN
(CHARLES, ALICE) : has-child
(CHARLES, CAR-1) : owns
(BETTY, CAR-2) : owns
CAR-1 : MERCEDES
CAR-2 : BMW

A visualization of the complete structure of the A-box is shown in Figure 2. If the agent is responsible for finding possible people, who most probably would buy certain types of cars, it might pose an instance retrieval query to the A-box of our host environment as follows.

concept-instances(MAN \sqcap (\exists *has-child* PARENT) \sqcap (\exists *owns* (BMW \sqcup MERCEDES)))

The result of this retrieval query is {CHARLES} because CHARLES is a man which has a child which, in turn, is a parent. Note that BETTY is a human because of the assumed range restriction for *has-child*. In addition, CHARLES owns a MERCEDES car. We assume that the individuals in the result set of this query may be candidates for advertisement operations. In

order to select appropriate material, we assume that the agent carries out additional tests as follows.

instance?(CHARLES, $(\exists \textit{has-child WOMAN} (\exists \textit{has-child WOMAN})) \sqcup$

$(\exists \textit{has-descendant} (\neg \textit{WOMAN} \sqcap (\exists \textit{owns BMW}))))$

Indeed, the answer is “yes”. However, it is not quite obvious why this is the case. Let us consider the first disjunct. From what is specified in our example A-box, CHARLES has a child which is a woman (ALICE) and who also has a child (BETTY). But for BETTY it is not known whether she is an instance of WOMAN. Let us assume that BETTY is indeed a woman. In this case the first disjunct is true and, therefore the answer is yes. If, on the other hand, BETTY is assumed to be an instance of $\neg \textit{WOMAN}$, the second disjunct becomes true because (implicitly) BETTY is a descendant of CHARLES, and BETTY owns a BMW.

Agents and Host Environment Use Different Ontologies

If an agent uses an ontology that is different from the ontology of a host environment visited by the agent, both ontologies must be related to each other. As an example we assume that the host environment uses the T-box defined above (let us call it T1), and the agent uses another T-box (T2). For instance, it might be the case that in T2 the agent uses concepts BMW_OWNER or MERCEDES_OWNER for a human who owns a BMW or Mercedes, respectively, rather than represent the same information with a role (relation) *owns* and an individual which is an instance of BMW or MERCEDES. The agent might then ask

concept-instances(MAN \sqcap $(\exists \textit{has-child PARENT}) \sqcap$

$(\textit{BMW_OWNER} \sqcup \textit{MERCEDES_OWNER}))$

If the query is answered w.r.t. the A-box of the host environment and its own T-box (T1), then the retrieval result is the empty set because the names BMW_OWNER or MERCEDES_OWNER are not related to the names used in T1. In T1 ownership is expressed with $(\exists \textit{owns} (\textit{BMW} \sqcup \textit{MERCEDES}))$. Hence, T1 uses a different style of conceptual modeling.

Since the host environment can check where the agent comes from, we assume that the host environment can involve a so-called ontology broker for providing inter-schema knowledge [8]. Inter-schema knowledge is related to different ontologies and can be conveniently expressed using terminological axioms in some description logic. We assume that an inter-schema ontology broker can supply the required axioms (possibly as a certain DAML+OIL resource that can be downloaded by the Racer Server of the host environment). The T-box T1 is extended by:

$\textit{BMW_OWNER} \equiv (\exists \textit{owns BMW}), \textit{MERCEDES_OWNER} \equiv (\exists \textit{owns MERCEDES})$

With these additional axioms, the agent gets the same answer as above. Due to space restrictions we cannot explain the use of inter-schema knowledge in detail and refer to [8] for further examples.

In a full-fledged information retrieval scenario, an agent might consult a document management system provided by an agent host environment. The agent can ask for documents that match a certain query in a similar way as discussed above. This scenario can also be realized with Racer if documents are annotated with metadata formalized with RDF. Information about documents can be represented using A-boxes (documents are represented by individuals) in the host environment. RDF annotations for documents are read by Racer and corresponding assertions are added to an A-box. Documents are represented as individuals. Agents can retrieve documents by posing retrieval queries to these A-boxes w.r.t. to specific T-boxes in the way exemplified above.

If we consider an instance retrieval query Q w.r.t. an A-box A , then it is clear that the solution set for Q could be extended if more information is added to A over time (whoever is responsible for that, another agent or the agent host environment). It would be a waste of resources to frequently poll the host environment with the same query (and repeated migration operations). Therefore, Racer supports the registration of queries at some server w.r.t. to some A-box (Publish/Subscribe Interface). With the registration, the agent specifies an IP address and a port number. The corresponding Racer Server passes a message to the agent if the solution set of a previously registered instance retrieval query is extended. The message specifies the new individuals found to be instances of the query concept Q . For details see the Racer manual [12].

Due to space restrictions, we can only give ideas for applications and services which can be implemented using logic. The examples we have given here should stimulate developers of agent systems and agent host environments to use the facilities of state of the art description logic inference engines. As a summary we discuss the features of the Racer System in the next section.

Features of the Racer System

Representation Language: Racer supports the language $\text{SHIQ}(\mathcal{D}_n)$ including reasoning about individuals (A-box reasoning). Large parts of DAML+OIL can be formalized using the description logic $\text{SHIQ}(\mathcal{D}_n)$. In the same way as in other DAML-OIL inference engines, individuals in concepts are approximated by Racer. For the time being there are some other minor restrictions on DAML+OIL and also RDF (see [12]). In particular, not all XML-Schema datatypes allowed in DAML+OIL are supported by Racer. Racer adopts the unique name assumption and the open-world assumption (for examples see the Racer manual [12]).

However, Racer also goes beyond DAML+OIL in many respects. Whereas DAML+OIL offers only data type reasoning based on subsumption between names, Racer also provides means for true algebraic reasoning on information about "concrete values" which are associated with individuals. Depending on the domain and the language for defining predicates over the domain, different so-called "concrete domains" are offered [11]. Racer supports interval reasoning over integers, inequations with order relations with respect to linear polynomials with real variables, and nonlinear multivariate non-linear polynomial (in)equations over complex numbers (Racer version 1-7-7).

Query Language: Racer supports multiple T-boxes and A-boxes and all query types introduced above.

Optimization Techniques: Various optimization techniques for ontology-based query answering with respect to T-boxes, A-boxes, and concrete values have been developed, implemented, and investigated with the Racer System. A few points are worth mentioning.

- *Language sensitivity:* One of the design goals of Racer is to automatically select state of the art optimization techniques that are applicable to the current input.
- *Optimized algebraic reasoning:* As a distinguishing feature Racer offers sound and complete algebraic reasoning as part of a logical inference engine. Optimization techniques for coupling algebraic and description logic reasoners provide for adequate system performance.
- *Query subsumption:* Instance retrieval queries can be answered in a faster way if the set of candidates can be reduced. In a similar way as for databases, the idea is to exploit results computed for previous instance retrieval queries by considering query subsumption (which is decidable in the case of the query language that Racer supports). However, this requires computing index structures for the T-box (the process is known as T-box classification) and, therefore, query subsumption is enabled on demand only.
- *Index generation on demand:* In some applications, A-boxes are generated on the fly with few queries referring to a single A-box. On the other hand, there are applications which pose many queries to more or less “static” T-boxes and A-boxes (which are maybe part of the agent host environment). The Racer Server supports both application scenarios. As a design decision, Racer computes answers for queries with as few resources as possible. Nevertheless, a Racer Server can be instructed to compute index structures in advance if appropriate to support multiple queries.

Persistency: In a similar way as in database systems, for query answering w.r.t. T-boxes and A-boxes complex data structures are computed and used internally by Racer. Internal structures of T-boxes and A-boxes being processed for query answering can be saved to disk for quick access and later reuse if the Racer Server is restarted.

Multi-User Support, Thread Safeness, Locking, Load Balancing: In a distributed systems context, there can be multiple agents connecting to a server at the same time. If they refer to the same A-boxes and T-boxes (which is very likely in the scenarios presented above), requests must be synchronized. Thus similar problems as with databases such as thread safeness, locking, and load balancing have to be dealt with. For instance, if multiple Racer Servers are started, queries can be automatically directed to “free” Racer Servers. These problems are tackled by the Racer Proxy, which is supplied as part of the Racer System distribution.

Conclusion

This article motivated what description logics can do for Semantic Web applications. We sketched the design of agent systems using practical description logic inference engines. In the context of the Semantic Web a dedicated architecture that is integrated into the distributed systems context is required. In particular, the article presented important features of the Racer System, and described its architecture and background. We motivated different kinds of queries that it supports and sketched possible scenarios for implementing logic-based services on the Semantic Web. Application demands and practical experiences with Racer demonstrated the importance of research on representation languages as well as query answering strategies to deal with realistic T-boxes and A-boxes that can be expected to be the

basis for implementing future agent systems operating on the Semantic Web (see [13, 14] for encouraging evaluation results).

References

- [1] F. Baader, Logic-Based Knowledge Representation. In M.J. Wooldridge and M. Veloso, editors, *Artificial Intelligence Today, Recent Trends and Developments*, number 1600 in Lecture Notes in Computer Science. Springer Verlag, 1999.
- [2] F. Baader, et al. (Editors), *The Description Logic Handbook: Theory, Implementation and Applications*, Cambridge University Press, 2003.
- [3] F. Baader, I. Horrocks, and U. Sattler. Description Logics as Ontology Languages for the Semantic Web, In Dieter Hutter and Werner Stephan, editors, *Festschrift in honor of Jörg Siekmann*, Lecture Notes in Artificial Intelligence. Springer-Verlag, 2003.
- [4] S. Bechhofer, I. Horrocks, C. Goble, R. Stevens, *OilEd: a Reason-able Ontology Editor for the Semantic Web*. Proceedings of KI2001, Joint German/Austrian conference on Artificial Intelligence, September 19-21, Vienna. Springer-Verlag LNAI Vol. 2174, 2001.
- [5] S. Bechhofer, The DIG Description Logic Interface DIG 1.1, <http://dl-web.man.ac.uk/dig/>.
- [6] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, May 2001.
- [7] D. Calvanese, D., M. Lenzerini, M., D. Nardi 1998. Description Logics for Conceptual Data Modeling. In Chomicki, J., and Saake, G., eds., *Logics for Databases and Information Systems*. Kluwer Academic Publisher.
- [8] D. Calvanese, Giuseppe De Giacomo, M. Lenzerini, Description Logics for Information Integration, In A. Kakas and F. Sadri, editors, *Computational Logic: Logic Programming and Beyond, Essays in Honour of Robert A. Kowalski*, volume 2408 of *Lecture Notes in Computer Science*. Springer, 2002.
- [9] R. Cornet, RICE is the Racer Interactive Client Environment, <http://www.b1g-systems.com/ronald/rice/>.
- [10] DAML+OIL (March 2001) <http://www.daml.org/2001/03/daml+oil-index.html>.
- [11] V. Haarslev, R. Möller, M. Wessel, The Description Logic ALCNHR+ Extended with Concrete Domains: A Practically Motivated Approach, In *Proc. International Joint Conference on Automated Reasoning, IJCAR'2001, June 18-23, 2001, Siena, Italy*.
- [12] V. Haarslev, R. Möller, Racer User's Guide and Manual, <http://www.fh-wedel.de/~mo/racer/racer-manual-1-7.pdf>.
- [13] V. Haarslev, R. Möller, High Performance Reasoning with Very Large Knowledge Bases: A Practical Case Study, In *Proc. 7th Int. Joint Conf. on Artificial Intelligence, IJCAI-01*, August 4-10, 2001, Seattle, Washington, USA.
- [14] V. Haarslev, R. Möller, Optimization Strategies for Instance Retrieval, In *Proc. Int. Workshop on Description Logics DL-2002*, Toulouse, 2002.
- [15] V. Haarslev, R. Möller, Racer System Description, In *Proc. International Joint Conference on Automated Reasoning, IJCAR'2001, June 18-23, 2001, Siena, Italy*.
- [16] I. Horrocks, U. Sattler, and S. Tobies. Reasoning with individuals for the description logic SHIQ. In David MacAllester, editor, *Proc. of the 17th Int. Conf. on Automated Deduction (CADE-17)*, number 1831 in *Lecture Notes in Artificial Intelligence*. Springer-Verlag, 2000.
- [17] O. Lassila and R. R. Swick. Resource Description Framework (RDF) Model and Syntax Specification. Recommendation, W3C, February 1999. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>.
- [18] S. McIlraith, T.C. Son and H. Zeng, Semantic Web Services, *IEEE Intelligent Systems, Special Issue on the Semantic Web*, Volume 16, No. 2, 2001.
- [19] OWL Web Ontology Language 1.0 Reference, <http://www.w3.org/TR/2002/WD-owl-ref-20020729/>.
- [20] RDFS, <http://www.w3.org/TR/2002/WD-rdf-schema-20020430/>.
- [21] S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, 1995.



Ralf Möller (Dr. rer. nat. habil.) is a Professor for Computer Science at the University of Applied Sciences in Wedel. His research interests include software technology for distributed systems as well as the application and theory of conceptual modeling and knowledge representation languages. His research goals encompass the development practical inference algorithms for embedding description logic systems into internet technology. He is a principal architect of the description logic reasoner RACER, which can be used as a core engine for building agent systems on the semantic web.



Volker Haarslev (Dr. rer. nat. habil.) is Associate Professor in the Computer Science Department at Concordia University in Montreal. He is internationally regarded for his substantial research contributions in the fields of visual language theory and description logics. His current research interests are in foundations of the semantic web and description logics, which play important roles in database and internet technology. He is a principal architect of the description logic reasoner RACER, which is a key component for the future of the semantic web.