Detecting mismatches between a user's and an expert's conceptualisations

Yongjian Huang, Vania Dimitrova, Pragya Agarwal

School of Computing, University of Leeds, UK
scs4yh@comp.leeds.ac.uk, vania@comp.leeds.ac.uk, pragya@comp.leeds.ac.uk

Abstract. The work presented in this paper is part of our ongoing research on applying commonsense reasoning to elicit and maintain models that represent users' conceptualisations. Such user models will enable taking into account the users' perspective of the world and will empower personalisation algorithms for the Semantic Web. A formal approach for detecting mismatches between a user's and an expert's conceptual model based on Description Logic is outlined. OWL-based rules are then derived and implemented in a demonstration prototype that is illustrated in a geographical domain using a SPACE ontology from the NASA's SWEET suite of ontologies for the Geo-spatial Semantic Web.

1 Introduction

Many web services are now being automated so that they can understand content, are able to process a vast amount of knowledge, and can provide more accurate and effective search of information resources [1]. The number of people who use these services is expanding, and hence, dealing with user diversity and providing personalisation functionality becomes paramount [2]. Users' preferences, expectations, goals and tasks differ. Moreover, people form *different conceptual models of the world* and these models dynamically change over time. On the other hand, knowledge-enhanced web services are normally driven by some description of the world which is encoded in the system in the form of an ontology defined by knowledge engineers[1]. The users' conceptualisation of the world may differ, sometimes significantly, from the conceptualisation encoded in the system. If not taken into account, the discrepancies between a user's and a system's conceptualisation may lead to the user's confusion and frustration when utilising Semantic Web services, which, in turn, can make these services less popular.

The long-term goal of our research is to apply commonsense reasoning approaches to capture and maintain users' conceptual models and to use these models for personalised, semantically-enhanced search on the web. For this, we consider that the domain expertise is encoded in some ontology (or several ontologies) pre-defined by domain experts and knowledge engineers. This expertise is used to guide the intelligent behaviour of the system and is combined with a model of the user that corresponds to the user's conceptualisation of the domain.

One possible way to elicit a users' conceptual model is by using interactive diagnostic agents [3, 4]. In an earlier work, conceptual graphs were employed to develop knowledge-enhanced user modelling techniques to elicit conceptual models of users [3]. Several levels of conceptual discrepancies were dealt with, such as misclassification, class misattribution, and individual misattribution, which triggered corresponding clarification dialogues with the user. However, the algorithms in [3] are based on conceptual graphs and are not interoperable. To enable sharing and interoperability, the expert ontology and the user's conceptualisation should be represented in a commonly accepted Semantic Web language, such as RDF or OWL. Denaux et al. [4] extended the interactive user modelling approach and implemented a diagnostic agent capable of extracting a user's conceptual model in OWL. The diagnostic agent was integrated in an adaptive learning content management system for the Semantic Web [4]. The work showed the strong potential for OWL-based interoperable user modelling approaches. However, although some discrepancies between a user's conceptualisation and the system's ontology were identified, there was no systematic study of what discrepancies may occur and the patterns were chosen rather ad-hoc.

This paper presents a systematic analysis of discrepancies that may occur between a user and a system's conceptualisation and proposes a formal approach, based on description logic, to define these patterns. OWL-based rules are then derived and implemented in a demonstration prototype that compares an expert ontology and a user's conceptualisation, both represented in OWL. The discrepancies are identified and registered as *misconceptions*.

2 Discrepancies between conceptualisations

Discrepancies between conceptualisations are inevitable in domains that are largely unstructured, such as the geographic domain. A standard terminology is not prevalent within such domains, and is dependent on the context of use and user. Indeterminacy and ambiguity in meanings are key issues in the development of ontologies in such domains. Smith [5] refers to this as the *Tower of Babel* problem where the heterogeneity in terminology and meanings leads to conceptual and terminological incompatibilities. Commonsense notions and cognitive conceptualisations structure knowledge by determining much of the conceptual and semantic content within the main categories. However, the inherent vagueness encapsulated in the terms and concepts leads to variability in the conceptualisations and causes conflicts and mismatches. Empirical results show that individual conceptualisations are characterised by semantic heterogeneity [6, 7], defined in [8] as *naming heterogeneity* (different names for identical concepts, characterised by synonyms and homonyms) and *cognitive heterogeneity* (different perceptions of real world phenomena).

Heterogeneity results in *conceptual discrepancies* (called also *mismatches*). Shaw and Gaines[9] point out that mismatches between concepts include using the same term for different concepts, using different terms for the same concept, or using different terms for different concepts. Wiederhold[10] extends these types to include mismatches at a deeper, semantic level, including: *Key difference* (different terms for the same concept), *Scope difference* (distinct domain coverage), *Abstraction grain* (varied granularity of detail among the definitions), *Temporal basis* (differences based on time categories), *Domain Semantics* (distinct domains), and *Value semantics* (differences in the encoding of values).

A more comprehensive list of discrepancies between two conceptual models is presented by Visser et al. [11] in their *explication mismatches* proposal which includes *concept definiens* and considers the following types of discrepancies: *Concept and Term mismatch* (same definiens but different concepts and terms); *Concept and Definiens mismatch* (same term but different concepts and definiens); *Concept mismatch* (same term and definiens but different concepts); *Term and Definiens mismatch* (same concept but different definiens and terms); Term mismatch (same concept and definiens but different terms); and *Definiens mismatch* (same concepts and terms but different definiens).

Algorithms for comparing two conceptualisations have been developed within ontology mapping research. Klein [12] summarises the main issues involved in combining ontologies, among which is the occurrence of *ontology level mismatches* which correspond to categories described above, such as domain coverage, concept scope, synonyms, homonyms, concept description, paradigm, and encoding. However, despite a considerable amount of work on ontology mapping, there is no formalisation available for comparing and finding semantic similarities between two conceptualisations.

To sum up, the review shows that while several researchers have attempted to describe mismatch patterns that may occur between two conceptualisations, the descriptions provided are vague at times and there is a lack of formal descriptions of these patterns. We will present below an initial formalisation that attempts to capture possible mismatches between a user's and an expert's conceptualisations. We aim to define a failry general framework that can incorporate most of the patterns discussed above.

3 Mismatch patterns defined with Description Logic

Formal approaches allow the design of algorithms at levels higher than the specific applications, and therefore, bring considerable insights into the design of intelligent system. In the next section, we will use Description Logic (DL) [13] to define discrepancies between a user's and a system's conceptualisations.

3.1 Illustration Domain

We will use examples from a geograph domain to illustrate the definitions below. This domain has been chosen because:

- it is a typical example of an *unstructured domain* with inherent vague concepts where users' conceptualisations may differ, see empirical studies in [6];
- geographical information systems are *widely used on the web*, and will be part of the next generation knowledge-based web;

– there is a *diversity of users* in this domain and many of them come from the general public;
– there *exist geographical domain ontologies*, for example the SWEET ontologies [1] from NASA, which is used in the illustrations below. The SPACE ontology that forms a part of the SWEET suite of ontologies is a good case example as SWEET (Semantic Web for Earth and Environmental Technology) is meant to be a foundation block for the Semantic Web by providing a common semantic interface for various web-based Environmental and Earth Sciences initiatives and is aimed at integrating various heterogeneous information sources.

3.2 Basic assumptions and notations

To avoid inconsistencies with the terminology (which differed across studies) we will follow a set of fairly conventional assumptions. *Concepts* are the main building blocks in a conceptual model and represent objects (either concrete or abstract) from the world. Each concept consists of four parts - *term*, *definiens* (also called definition), and *properties* - which define the *intentional meaning* of a concept. We will first reason with definiens of two concepts for their set relationship, and then check the term and property for misconception. If all three parts of two concepts match, we consider that there is no misconception between these two concepts. The following notations will be used:

C_u: the concept in the user's conceptualisation.
C_e: the concept in the expert's conceptualisation.
I_u: the individual of a concept in the user's conceptualisation.
I_e: the individual of a concept in the expert's conceptualisation.
P_u: the property identified in the user's conceptualisation.
P_e: the property identified in the expert's conceptualisation.
Term(C): the name of the concept
$\forall R \cdot E(C)$: value restriction for the concept C.
$=$: equality, as owl:sameAs
$\equiv$: equivalence, as owl:equivalentClass
$\sqsubseteq$: subsumption, as rdfs:subClassOf
$\sqsubseteq_D$: one concept is a parent of another,as directly rdfs:subClassOf in OWL
$\Cap$: partial overlap
$\circ$: disjointness

The mismatch patterns will be defined as rules. The right hand side of $\vdash$ is the relationship between two concepts. The left hand side of $\vdash$ is the original information available to the proof system. We define *five set relationships* between two concepts, namely, *equality*, *equivalence*, *subsumption*, *partial overlap*, and *disjointness*.

[1] http://sweet.jpl.nasa.gov/ontology/

3.3 Mismatches based on equality

Equality indicates that two concepts have exactly the same intentional meaning. Given the proof $C_u =_{def} D, C_e =_{def} D \vdash C_u = C_e$ where D can be either atomic concept or combination of other concepts, we define:

Term Mismatch $Term(C_u) \neq Term(C_e) \rightarrow$ Term Mismatch

This is the case when two concepts share the same intentional meaning but have different names. For example, in a user's conceptualisation, there is *Floor* owl:sameAs *Base* whereas in an expert's ontology there could be *Bottom* owl:sameAs *Base*. Thus, *Floor* is equal to *Bottom* even though they have completely different names.

Attribute Mismatch $\forall R \cdot E(C_u) \neq \forall R \cdot E(C_e) \rightarrow AttributeMismatch$

This is the case when two concepts share the same intentional meaning but have different attributes. For example, the user can say a *region* is *above Y-coordinate*, while the expert ontology says a *region* is *over Y-coordinate*.

3.4 Mismatches based on equivalence

Two concept are considered equivalent if they have the same set of individuals, i.e. $C_u \equiv_{def} D, C_e \equiv_{def} D \vdash C_u \equiv C_e$ where D can be either atomic concept or combination of other concepts.

Term Mismatch $Term(C_u) \neq Term(C_e) \rightarrow TermMismatch$

Two concepts have the same sets of individuals, however, the concepts may have different intentional meaning. There are many examples in the space ontology. For instance, *edge* is equivalent to *boundary*, yet these two concepts have different intentional meanings in their own rights.

Attribute Mismatch $C_u \equiv_{def} D \sqcap \forall R \cdot E, C_e \equiv_{def} F \sqcap G, \vdash R_u \equiv R_e$ where D, F, G can be any concepts.

$F \sqsubseteq \forall R \cdot E, Term(C_u) = Term(C_e) \rightarrow AttributeMismatch$

This is so-called attribute assignment mismatch[11], which is a property misconception occurring when two properties are the same except the domains, with one being a subset of another. For instance, a user assigns to *HorizontalCoordinate* the attribute of *hasDirection*, which could be assigned to *Coordinate* in the expert ontology.

Value Mismatch $C_u \equiv_{def} D \sqcap \forall R \cdot E, C_e \equiv_{def} F \sqcap \forall R \cdot G \vdash R_u \equiv R_e$ where D, F, G can be any concepts.

$Term(E) \neq Term(G), Term(C_u) = Term(C_e) \rightarrow ValueMismatch$

Similar to the previous pattern, the two properties have everything equivalent except the range.

Individual Mismatch $C_u \equiv_{def} D, C_e \equiv_{def} D \vdash C_u \equiv C_e$ where D can be either atomic concept or combination of other concepts

$C_u(I_u), C_e(I_e), \forall i, Term(I_{ui}) \neq Term(I_e) \rightarrow IndividualMismatch$

Individuals of the same class may have different names, but they should not affect the equivalence of two concepts with the same extension. For example, for the concept *longitude*, one conceptualisation assumes its instances to be a number of *miles* whereas another conceptualisation with the same class assumes its instances to be a number of *kilometers*.

Abstraction Mismatch $C_u \equiv_{def} (D_1 \sqcup D_2 \sqcup \cdots \sqcup D_m), C_e \equiv_{def} (D_1 \sqcup D_2 \sqcup \cdots \sqcup D_m) \vdash C_u \equiv C_e$ where D_i can be either atomic concept or combination of other concepts

C_u does not exists $\rightarrow$ Abstraction Mismatch

This mismatch occurs when user has a concept whose abstraction does not exist in expert ontology. For example, *Coordinate* usually include *HorizontalCoordinate* and *VerticalCoordinate*, but *Coordinate* could be missing in the user's conceptualisation, that is, the user is not aware that both of coordinates form the whole coordinate for a location.

3.5 Misconceptions based on subsumption

Subsumption shows that one concept is a sub-class of another.

Structure Mismatch $C_u \equiv_{def} (D_1 \sqcap D_2 \sqcap \cdots \sqcap D_m \sqcap \forall R_1 \cdot F_1 \sqcap \forall R_2 \cdot F_2 \sqcap \cdots \sqcap \forall R_k \cdot F_k), C_e \equiv_{def} (E_1 \sqcap E_2 \sqcap \cdots \sqcap E_n \sqcap \forall S_1 \cdot G_1 \sqcap \forall S_2 \cdot G_2 \sqcap \cdots \sqcap \forall S_l \cdot G_l) \vdash C_u \sqsubseteq C_e$ where D_i, E_i can be either atomic concept or union of other concepts

$\forall i, 1 \leq i \leq m, \exists j, 1 \leq j \leq n, D_i \sqsubseteq E_j$, and
$\forall i, 1 \leq i \leq k, \exists j, 1 \leq j \leq l, R_i = S_j, F_i \sqsubseteq G_j$

$Term(C_u) = Term(C_e) \rightarrow StructureMismatch$

The description is similar to a subsumption problem. The only difference is the last condition, which indicates concept subsumption with structure mismatch or Definiens Mismatch. For instance, the user may define *Top* as *maximal height*(*Top* $\equiv_{def}$ $\forall$*hasHeight* $\cdot$ *Maximum*) whereas expert ontology defines *Top* as *Maximum* with *up direction*(*Top* $\equiv_{def}$ *Maximum* $\sqcap$ $\forall$*hasDirection* $\cdot$ *Up*).

3.6 Misconceptions between partially overlapping concepts

Scope Mismatch $C_u \equiv_{def} (C \sqcap D), C_e \equiv_{def} (E \sqcap F) \vdash C_u \cap C_e$ where C, D, E, F can be atomic concepts or complex concepts; C, E can also be bottom concepts; and $(C \sqcap E) \sqsubseteq \perp, D \cap F$

$Term(C_u) = Term(C_e) \rightarrow ScopeMismatch$

Note that this is a recursive definition. If two concepts share most of instances, then they overlap with one another. There could be a possible misconception between two equivalent concepts.

3.7 Misconceptions between Disjoint Concepts

Definiens Mismatch $C_u \equiv_{def} (C \sqcap D), C_e \equiv_{def} (E \sqcap F) \vdash C_u \circ C_e$ where C, D, E, F can be atomic concepts or complex concepts; C, E can also be bottom concepts; and $(C \sqcap E) \sqsubseteq \perp, D \circ F$

$Term(C_u) = Term(C_e) \rightarrow DefiniensMismatch$

This could be a huge misconception, in that two concepts in the same name have completely different definiens. The mismatch reflects that the user's view of a concept is totally different from the expert's point of view. For example, the user may consider *profile* as *outline*, but *profile* is defined as *horizon* in the expert ontology.

4 Implementation

Because the misconception patterns have been defined in Description Logic, they can be applied to conceptualisations defined in OWL. The ontology can be taken from any domain. In the current trials, we have used the NASA SPACE ontology that incoporates concepts from the Earth and Environment terminology. The illustrations in Section 3 were based on this ontology. A snapshot of this ontology, used as the expert ontology in the context of our work, is given in figure **??**.

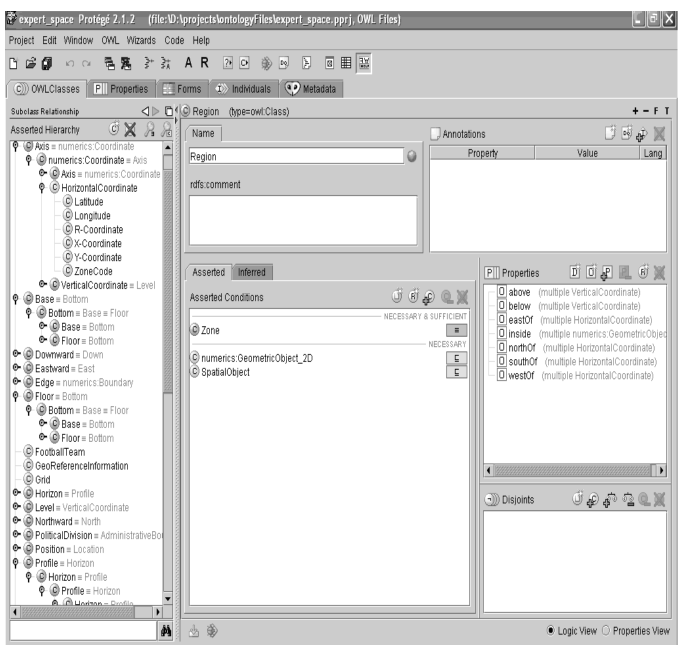

Fig. 1. An extract from the NASA space ontology, shown in Protege.

Based on the formal descriptions of mismatches, we have implemented algorithms to capture a user's misconceptions defined as the discrepancies between the user's and the expert's perspective of the world. For this we have used the rule-based OWL inference engine Jena2[14].[2] The five types of relationship are defied in OWL, as follows

```
owl:same      equality
owl:eq        equivalence
owl:subsume   subsumption
owl:po        partial overlap
owl:dj        disjointness
```

Following the triple-based nature of Jena, we have defined a set of rules to capture the mismatches defined in Section 3. For example, to check for *Structure Mismatch* based on *subsumption* relation between two concepts, the following rule will be passed to Jena:

```
[pattern8: (?C owl:and ?X), (?C owl:and ?Y), (?D owl:and ?U), (?D
owl:and ?V) , (?X owl:eq ?U), (?Y owl:eq ?V), notEqual(?C, ?D),
notEqual(?X, ?Y), notEqual(?U, ?V) -> (?C owl:eq ?D)]
```

To illustrate, let us consider two concepts from the user's and the expert's conceptualisations, respectively. Let both concepts have the same term that is *HorizontalCoordinate*. In the expert's ontology *HorizontalCoordinate* is defined as *Coordinate* and *hasNormal Up* (see Figure 1), whilst in the user's conceptual model, it is defined as *Axis* and *hasNormal Up*. However, *Axis* and *Coordinate* can be proved to be equivalent. Hence, the two concepts are still equivalent but with structure mismatch. Figure 2 shows the interface of a Java application we have implemented to test the misconception detection patterns. In this application, both the user's and the expert's conceptualisations are passed as OWL-models (i.e. a collection of OWL statements). The figure shows that the algorithm identifies the pattern of ontology misconceptions and compares the two concepts considering their *neighbourhoods*, i.e. their terms, properties, children, and parents.

5 Current state and future work

The work presented in this paper is part of our ongoing research on applying commonsense reasoning to elicit and maintain models that represent users' conceptualisations of the real world. Such user models will enable taking into account the users' perspective of the world and will empower personalisation algorithms for the Semantic Web. A formal approach for detecting mismatches between a user's and an expert's conceptual model is outlined. The formalisation is used as the basis to develop algorithms to compare two conceptualisations defined in OWL. The algorithms are illustrated in a geographical domain using a SPACE

[2] Jena can be downloaded from http://jena.sourceforge.net, and a comprehensive documentation is available there.

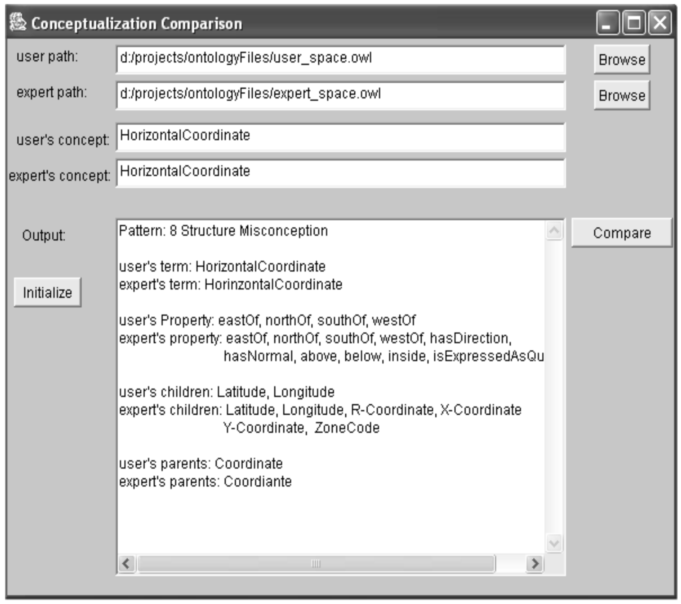

Fig. 2. A screen shot of the Java prototype for misconception detection.

ontology developed as part of the SWEET initiative for the Semantic Web by NASA, and have been tested by simulating possible user misconceptions.

Our immediate plans are to conduct more appropriate evaluation of the algorithms based on real data of possible user's conceptualisations. Because we experiment with a fairly intuitive domain of earth sciences categories, it is feasible to conduct studies where people (naive domain users) are asked to identify the main concepts and their relations, on the basis of which user conceptual models can be extracted, see for example [6]. These conceptualisations may be used to test and fine-tune the algorithms.

There are three possible Semantic Web applications of the algorithms presented in this paper. Firstly, algorithms that highlight discrepancies between a user's conceptualisation and an existing ontology can be used to *fine-tune the expert ontology.* We are also looking into the possibilities of using the mismatch detection algorithms in combination with some additional reasoning to deal with vagueness and heterogeneity problems, as discussed in section 2. Secondly, a systematic approach for detecting mismatches can be combined with dialogue planning, such as OWL-OLM [4], to develop robust interactive diagnostic agents that *extract and validate a conceptual model of the user.* Finally, any user modelling approaches that *take into account different viewpoints of the world* can benefit from the algorithms presented here. It is worth noting that although the conceptualisation discrepancies have been called misconceptions in this paper (for consistency with the user modelling terminology), these discrepancies are by no means a sign of erroneous beliefs. The Semantic Web paradigm and the open world assumption requires the deployment of appropriate user modelling approaches that capture and maintain different user perspectives.

Acknowledgments. The authors thank Ronald Denaux and Michael Pye whose Java implementation of OWL-OLM was helpful for the set up of the testing environment. Many thanks to Lora Aroyo for stimulating discussions, and the anonymous reviewers whose comments helped improve the quality of this paper.

References

1. Ding, Y., Fensel, D., Klein, M.C.A., Omelayenko, B.: The semantic web: yet another hip? Data Knowledge Engineering **41** (2002) 205–227
2. Henze, N.: Personalization functionality for the semantic web: Identification and description of techniques. Technical report, REWERSE project: Reasoning on the Web with Rules and Semantics (2004)
3. Dimitrova, V., Bontcheva, K.: Knowledge-based approaches for interactive student modelling and adaptive web explanations. In: Proceedings of the 7th international conference on Knowledge-Based Intelligent Information and Engineering Systems. (2003) 230–237
4. Denaux, R., Dimitrova, V., Aroyo, L.: Integrating open user modeling and learning content management for the semantic web. In: Proceedings of UM2005. (2005)
5. Smith, J.M.: Large-scale knowledge systems. Foundations of knowledge base management (1989) 259–281
6. Agarwal, P.: Contested nature of 'place': knowledge mapping for resolving ontological distinctions between geographical concepts. In Egenhofer, Freksa, Miller, eds.: Geographic Information Science, Proceedings of 3rd International Conference, GIScience 2004, Berlin, Springer-Verlag (2003) 1–21
7. Hameed, A., Sleeman, D., Preece, A.: Detecting mismatches among experts' ontologies acquired through knowledge elicitation. Knowledge-Based Systems **15** (2002) 265–273
8. Bishr, Y.: Overcoming the semantic and other barriers to gis interoperability. International Journal of Geographical Information Science **12(4)** (1998) 299–314
9. Shaw, M., Gaines, B.: Comparing conceptual structures: Consensus, conflict, correspondence and contrast (1989)
10. Wiederhold, G.: Mediators in the architecture of future information systems. In Huhns, M.N., Singh, M.P., eds.: Readings in Agents. Morgan Kaufmann, San Francisco, CA, USA (1997) 185–196
11. Visser, P.R.S., Jones, D.M., Bench-Capon, T.J.M., Shave, M.J.R.: An analysis of ontological mismatches: Heterogeneity versus interoperability. In: AAAI 1997 Spring Symposium on Ontological Engineering, Stanford, USA (1997)
12. Klein, M.: Combining and relating ontologies: an analysis of problems and solutions. In: IJCAI-2001 Workshop on ontologies and information sharing. (2004) 53–62
13. Baader, F., Nutt, W.: Basic description logic. In Baader, F., Calvaness, D., McGuinness, D., Nardi, D., Patel-Schneider, P., eds.: Description Logic Handbook. Cambridge University Press (2002) 47–100
14. Carroll, J.J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., Wilkinson, K.: Jena: implementing the semantic web recommendations. In: Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, ACM Press (2004) 74–83