BestMap: Context-Aware SKOS Vocabulary Mappings in OWL 2

Rinke Hoekstra[1,2]

[1] Department of Computer Science, Vrije Universiteit Amsterdam
hoekstra@few.vu.nl
[2] Leibniz Center for Law, Universiteit van Amsterdam
hoekstra@uva.nl

Abstract. This paper describes an approach to SKOS vocabulary mapping that takes into account the context in which vocabulary terms are used in annotations. The standard vocabulary mapping properties in SKOS only allow for binary mappings between concepts. In the BestMap ontology, annotated resources are the contexts in which annotations coincide and allow for a more fine grained control over when mappings hold. A mapping between two vocabularies is defined as a class that groups descriptions of a resource. We use the OWL 2 features for property *chains*, *disjoint* properties, *union*, *intersection* and *negation* together with careful use of equivalence and subsumption to specify these mappings.

1 Introduction

Vocabulary mapping is an important task in many domains. Most notably the field of cultural heritage – with libraries and museums being its prime representatives – has a lot to gain from these mappings. Because these information providers have already been using highly elaborate, rich metadata schemes and thesauri for many decades, a mapping between these schemes allows the uniform disclosure of information coming from large heterogeneous data sources. It is not surprising, therefore, that these fields have developed their own mapping and alignment technology, prior to, and independent from efforts within the Semantic Web community (cf. [1] for an overview).

Such Knowledge Organisation Systems (KOS) provide primitives for the systematised annotation of large volumes of resources. This is to guarantee retrievability of these resources. The purpose of KOS is therefore to enable *concept-based information retrieval*, rather than reasoning services. The Simple Knowledge Organisation System (SKOS) is an attempt to integrate the expertise in these fields with existing Web standards such as RDF Schema [2].[3] SKOS leverages existing metadata by providing language constructs that are commonly used in library science, without imposing more expressive semantics. This is probably the most distinguishing feature of SKOS in relation to OWL 2 [3].[4]

[3] SKOS: developed by the W3C (Recommendation), see `http://www.w3.org/TR/skos-reference/`.
[4] See `http://www.w3.org/TR/owl2-overview/`.

A consequence of the design of SKOS is that vocabulary terms – 'concepts' – in SKOS are *instances* of the skos:Concept class.[5] Relationships between concepts are captured using e.g. skos:narrower, skos:broader and skos:related, rather than rdfs:subClassOf. This allows the representation of both 'isa'-like and 'part-of' relations using the same relation, as is common in many thesauri. As any SKOS vocabulary is 'just' RDF, the set of standard relations can be extended using the rdfs:subPropertyOf predicate.

SKOS offers several properties for specifying *mappings* between concepts in different vocabularies – or 'concept schemes'. Mappings allow the traversal of multiple vocabularies for the purposes of information retrieval in heterogeneous information sources (cf. [4] and many others). The inventory of [1] shows that the types of mapping relations used in practice is rather limited. And again, it is the practice of library science and cultural heritage that formed the use case for vocabulary mapping relations in SKOS.

The following section sketches a use case for contextual many-to-many mappings (section 2). Section 3 gives an overview of the limitations of the vocabulary mappings supported by SKOS. Section 4 lists the requirements for vocabulary mappings in the BestPortal [5]. We present the BestMap ontology that implements an approach that does allow for expressive mappings using OWL 2 DL in section 5.

2 Use Case: The BEST Project

The BEST project, described in e.g. [6], is a research project that aims to improve the accessibility of court proceedings related to tort law, which are collected and published by the Netherlands Council of the Judiciary.[6] This is done first by developing a portal [5] that allows the "Man on the Clapham omnibus" – our typical reasonably educated layman – to find relevant case law, and assess whether his case has any chance in court.[7] In other words, it enables him to become aware of his best alternative to a negotiated agreement, his BATNA. The main reason for going to court is to attain a better situation than that which would otherwise come about. If you are not aware of this 'status quo' situation, you run the risk of either achieving an outcome that you should have avoided, or missing out on an outcome that would have left you better off.

A central issue in legal information systems is the peculiar and precise language used by lawyers. This *legalese* is hard to understand, easy to misinterpret and often plain incomprehensible for laymen. Such a system therefore has to

[5] Note to the reader: skos:Concept is an owl:Class but an instance of skos:Concept may also be an instance of owl:ObjectProperty, or of owl:Class itself (both valid in OWL 2 through punning).

[6] BEST: BATNA Establishment using Semantic Web Technology, see `http://www.best-project.nl`

[7] Officially, the Man on the Clapham omnibus is 'the reasonable person', a hypothetical person against whom a defendant's conduct may be judged in an English law civil action for negligence.

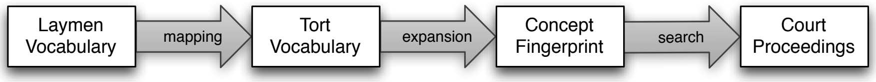

Fig. 1. Accessing court proceedings in the BEST portal.

somehow 'bridge the gap' between legal knowledge and commonsense knowledge [7]. However, the possibility of the legal qualification of a case – by legal professionals – is a strong indication that the vocabularies of legal knowledge and common sense are not disjoint. The assumption that abstract legal terminology can be reduced to the actual societal events and states governed by law [8], is a key aspect of e.g. the LKIF Core ontology of [9,10] and the approach to normative assessment of [11].

However, knowledge-based legal information systems rely on a strong, formal connection between the two domains. Rather than a mere mapping between terminologies, this connection requires the *interpretation* of a case in legal terms: the rigorous definition of legal concepts in terms of common sense notions. Such a scenario is not realistic for a system that is intended to give users a 'feel' rather than a verdict over their situation. Furthermore, the current database of the Council of the Judiciary serves some 50 thousand preselected court proceedings.[8] Taking a knowledge-based approach would either require the representation of every single one of them, or a much smaller selection of court proceedings – inevitably leaving out some of the more interesting, 'special' cases.

The approach we take in BEST is that of concept-based search: users can specify their case using the commonsense concepts of their choosing. The system maps these to corresponding legal concepts, i.e. it rephrases the search concepts using legal concepts, after which these concepts are expanded to a weighted list of literal terms – the fingerprint – used to search in the case database (see Figure 1).

A good question to ask, at this point, is: why go through all this trouble? The approach described here requires the construction of a layman vocabulary, a tort vocabulary and a mapping between the two. How is this different from a more rigorous knowledge based approach? Wouldn't a simple full text search as is already offered by the Council be more effective?

Usefulness of full text search is held back by the way in which court proceedings are written down. This holds both for laymen and for lawyers. The layman will simply be unable to phrase his query in such a way that relevant documents are retrieved. For instance, a search on *horse*, *riding school* and *kick* may retrieve the one case in which a horse kicked someone at a riding school, but not the plethora of other cases in which the owner of an animal can be held

[8] The database is disclosed to the general public via `http://www.rechtspraak.nl`. This website currently offers search on several fields, including full text search on the text of court proceedings.

liable for any damages caused to a third party. The lawyer, on the other hand, will not only be subject to the general limitation of full text search – a concept may manifest itself as many literal strings – but will also have to deal with the disparateness of terminology between the written law – legislation – that defines a legal concept and the case law that applies it [6]. If he is lucky, the court proceeding will make reference to the applicable articles.

3 Limitations of Standard Mappings

The SKOS reference defines six mapping relations: *mappingRelation*. *narrowMatch*, *broadMatch*, *relatedMatch*, *closeMatch*, *exactMatch*. [9] These relations are used to define mappings between pairs of skos:Concept individuals that (usually) reside in two different skos:ConceptSchemes. In this section we briefly explore the limitations of SKOS mapping relations in a broader context.

Recent vocabulary mapping efforts are directed at resolving two issues, *representational* and *conceptual* heterogeneity [4], respectively: the construction of a SKOS-compatible vocabulary, given a terminology in some legacy format [12], and the discovery of vocabulary mapping relations, given two vocabularies (lexical or structure based), or two vocabularies and a set of annotated resources (instance-based, cf. [13]). To confuse matters, both approaches define a *mapping* of some sort. In the first approach, the mapping is between the legacy terminology and a SKOS vocabulary. Here, mapping is the construction of a set of *rewriting* rules, that perform operations such as string concatenation etc. [12]. These rules are usually expressed as XSLT transformations, or code in e.g. Java. For the second case the mappings are much more straightforward, as the vocabularies involved are already expressed in RDF. Mappings are commonly stated using the standard set of relations defined in SKOS.

The SPARQL++ extension [14] to the SPARQL language [15], is meant to accomodate several common syntactic incompatibilities between RDF vocabularies that require non-trivial mappings. This solution is suitable only in cases involving the *transformation* of information stated using one vocabulary to that of another, and not for simultaneous disclosure of heterogeneous information via multiple vocabularies. Other surveys indicate that also at the *conceptual* level, the field may well benefit from more expressive mapping rules. Isaac et al. [16] show that existing tools do not produce hierarchic nor many to many mappings, which – in their view – is due to limited "knowledge of realistic application domains" [16, p.2]. Many-to-many mappings are a requirement for multi-vocabulary *search* across collections, and the *reindexing* of a collection with a new vocabulary.

SKOS meets the requirement of hierarchic relations using the skos:broadMatch and skos:narrowMatch mapping relations, though cardinality is only minimally supported. Multiple skos:broadMatch assertions can express multiple one to many mappings for a single concept. Many to many mappings can only be simulated using multiple one to one mappings.

[9] See `http://www.w3.org/TR/skos-reference/`.

3.1 Intensionality vs. Extensionality

The problems described above are mainly due to the fact that the semantics of SKOS concepts is intensional. That is, the standard interpretation of a concept is the concept itself, not its instances. Relations, therefore, are intensional as well, and do not necessarily impose restrictions on the use of concepts involved. This perspective is also present in the mapping properties of SKOS.

The information retrieval perspective that is the driving force behind knowledge organisation systems and the intensional relations of SKOS seem to be at odds. Lightweight semantics are closely tied to the information retrieval perspective, but this is not the case for the intensionality of SKOS relations. This may not be a problem where it concerns direct relations between SKOS concepts within a single concept scheme: relations are either straightforward (skos:broader and skos:narrower) or very weak (skos:related). However, an underlying assumption is that resources *annotated* by some concept should also be *retrievable* via its broader concept: there is an implicit *extensional* assumption.[10]

This assumption of extensionality is most prevalent in cases where different vocabularies need to be aligned, and collections reindexed to allow for integrated access to heterogeneous collections [4]. In fact, evaluation of the quality of alignment can only realistically be done by taking the extension of concepts into account [16]. This is also the main motivation for the instance-based vocabulary mapping approaches mentioned in section 3. Here, a mapping relation is constructed on the basis of a measure of correspondence between the extension of concepts in either vocabulary.

4 Requirements

Mappings between layman-used and legal terms in the BEST portal have to meet the following requirements. Ideally, relations are specified using pre-existing SKOS mapping relations: the mappings should be SKOS-*compatible*. An additional mapping mechanism may exist, but should not interfere with SKOS. Secondly, the mappings should not be too rigid, and take into account that a mapping is *not a definition*, but rather a means to retrieve (possibly) relevant court proceedings from a case database.

Thirdly, there should be support for many-to-many mappings. In [16], many to many mappings are deemed necessary primarily for reasons of granularity: one vocabulary may make distinctions that another does not. In our case, granularity is of secondary importance, but it is rather the *context* in which layman concepts co-occur in the annotation of an individual case that determines whether a legal concept is part of a valid characterisation of that case. Considering each concept in the case of the kicking horse separately is less informative than considering them in meaningful groups. For instance, a *riding school* can play many roles, as a building that has burnt down, the duped buyer of fodder in a sales transaction,

[10] The occurrence of an individual skos:Concept in the annotation of a resource is part of the extension of the concept.

the location of a crime, etc. However, in our case, the school is most relevant as the *owner* of an *animal* (the horse) that caused some *damage* (the kick). If the person being kicked is an employee, then it might also be a case of liability for *subordinates*.

Lastly, specifying mappings at this level of specificity is not very flexible, and we should therefore exploit the skos:narrower and skos:broader hierarchy to allow for similar cases to be covered by a single mapping. Section 5 introduces a framework, or design pattern if you will, for specifying context-aware mappings between two (or more) vocabularies, and specifies means to extend this framework by including more contextual information using OWL 2 role chains.

5 BestMap: Context-Aware Mappings

As has been made clear in the previous sections, mapping relations often rely on intensional semantics, where the applicability of mappings is often existentially defined. In our approach, we exploit the power of OWL 2 DL to define restrictions on the members of sets of annotated resources as intensional descriptions. The BestMap ontology makes explicit the medium – a resource – through which a mapping can be specified.[11] To do this, we introduce the property :about and its inverse that allow us to annotate resources with concepts from a SKOS vocabulary:

$$\text{:about} \equiv inv(\text{:describes})$$

The :about property relates a :Resource (its domain) with some skos:Concept (its range). The reason for introducing a new relation is that other options, such as *punning* of SKOS concept individuals, or reusing existing vocabulary are limited in the sense that they cannot be subject to OWL 2 class restrictions.

As we are dealing with SKOS vocabularies, it seems useful to integrate SKOS's semantic relations with this approach. There are two issues here. First, simply defining a sub property role chain that reuses :about is not allowed in OWL 2 as it introduces a cycle [3]. We therefore need to distinguish between *direct* annotations, and *indirect* annotations:

$$\begin{aligned} \text{:d\_about} &\sqsubseteq \text{:about} \\ \text{:d\_describes} &\sqsubseteq \text{:describes} \\ \text{:d\_describes} &\equiv inv(\text{:d\_about}) \end{aligned}$$

Secondly, SKOS makes a distinction between quite a variety of semantic relations. We define sub properties of :about and :describes for each of these, taking into account the direction of the relation. For instance:

$$\begin{aligned} \text{:d\_about } \mathbf{o} \text{ skos:broaderTransitive} &\sqsubseteq \text{:bt\_about} \\ \text{:d\_about } \mathbf{o} \text{ skos:narrowerTransitive} &\sqsubseteq \text{:nt\_about} \end{aligned}$$

[11] The BestMap ontology is available from `http://www.best-project.nl/owl/bestmap.owl`.

The resulting property tree for :about is depicted in Figure 2. We assert inverses and property disjointness where appropriate. Clearly, :bt_about and :nt_about are disjoint. However, :d_about is not disjoint with any of the other properties as a resource may be annotated by two concepts that are related via one of the SKOS properties.

The sub properties of :describes are simply defined as the inverse of the corresponding property in the :about hierarchy:

$$:\mathsf{r\_about} \equiv inv(:\mathsf{r\_describes})$$
$$:\mathsf{bt\_about} \equiv inv(:\mathsf{nt\_describes})$$

The property assertions allow the inference that any resource directly annotated with a SKOS concept using the :d_about property is also annotated by the concepts related to it via standard SKOS relations, coming from any mapped vocabulary. A front-end that uses SPARQL queries to retrieve annotations can readily make use of either :about or its sub properties to take into account the variability in annotation strength.

Note that because these properties are defined using sub property chains, they are not a shorthand for the chain and should never be used directly to annotate resources: the existence of a property assertion does not guarantee the existence of the property chain.[12] OWL 2 does not provide a notion that can be used to preclude assertions using these inferable properties. Furthermore, OWL 2 Full does not allow us to express equivalence between a property and a chain.[13]

5.1 Mapping Classes

It is now straightforward to identify classes of resources by specifying class axioms that restrict the possible values of these mapping properties. A mapping relation is then an OWL class that defines restrictions on these properties, using concepts from two vocabularies. No generic mechanism exists in OWL 2 that allows us to express that the two vocabularies should be distinct. However, it is not too difficult to imagine scenarios where complex relations between concepts from a single vocabulary can be useful.

Since mappings are just OWL classes, we can use any OWL 2 construct of our choosing to define them. Furthermore, mappings can be *directed* in the sense that annotations using concepts from one vocabulary will lead to inferred annotations using concepts of the other vocabulary, but not the other way around.

For instance, in the case of the riding school we can define a mapping that translates from the generic layman vocabulary to the legal tort vocabulary as

[12] The same holds for the :about and :describes property, their direct variants should be used instead.

[13] The structural specification of OWL 2 [17] distinguishes `subObjectPropertyExpression` from `superObjectPropertyExpression`, the latter of which cannot contain a `propertyExpressionChain`. This means we cannot express equivalence between a property and a chain by simply asserting two `SubObjectPropertyOf` statements with reversed arguments.

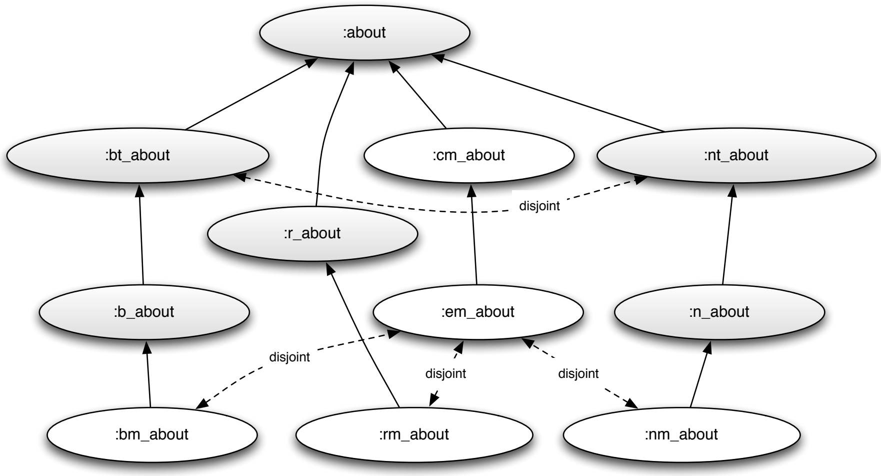

Fig. 2. Redefinition of SKOS relations using :about.

follows:

$$\begin{aligned}\textsf{ex:AO\_Mapping} \equiv\ & \textsf{:about } \textbf{value } \textit{lv:animal} \sqcap \textsf{ :about } \textbf{value } \textit{lv:company} \sqcap \\ & \textsf{:about } \textbf{value } \textit{lv:dangerous\_action} \\ \sqsubseteq\ & \textsf{:d\_about } \textbf{value } \textit{tv:animal} \sqcap \textsf{ :d\_about } \textbf{value } \textit{tv:animal\_owner} \sqcap \\ & \textsf{:d\_about } \textbf{value } \textit{tv:damage}\end{aligned}$$

Admittedly, this mapping is quite unrestricted as it triggers even for resources annotated with concepts that only have a skos:related relation with the concepts of the restriction. It is easy to see how the more specific relations of Figure 2 can be used in the class axiom to further restrict the applicability of the mapping.

The mapping is *directed* since a resource annotated with the legal concepts of ex:AO_Mapping will not be classified as an instance of that class. The direction can be reversed by swapping the subclass and equivalent class restrictions on annotations. Furthermore, two annotations can be defined as equivalent by using only equivalent class restrictions. There is of course no restriction on the use of additional OWL 2 expressions in the class axiom. However, for the mapping it makes no real sense to use a union in the equivalent class restriction: this will not lead to new (known) :d_about relations between matching resources and concepts. One step further is the possibility to *exclude* resources from a mapping where they are annotated using some concept. This can be done by introducing the complement of a value restriction in the equivalent class axiom of a mapping. However, because of the open world assumption this does not mean that resources that are not annotated will be members of the mapping class: this requires explicit negative property assertions on the :d_about property (thankfully available in OWL 2).

6 Discussion

This paper presents the BestMap ontology for specifying context-aware SKOS vocabulary mappings. It introduces the use case that formed the basis of the ontology in section 2: a portal meant to improve access to court proceedings for people without a legal background. This portal relies on a flexible mapping between layman concepts and legal terms. The main requirement for this feature is many to many mappings between these concepts. We investigated the mapping relations provided by SKOS, and discussed several shortcomings (section 3). One of these is that in SKOS, mappings between concepts are purely intensional: there is no connection to the resource annotated by the concepts. This seems strange as knowledge organisation systems are primarily targeted to the disclosure of information, and mappings between vocabularies can therefore only be evaluated by reference to the quality of annotations produced by the mapping.

BestMap introduces the :about property to annotate resources. Sub properties of this property are defined as role chains that include SKOS relations. A mapping is a class that uses value restrictions on the :about property or one of its sub properties in an equivalent class axiom to classify resources annotated using the d_about property. The result of this classification is that additional annotations are inferred for the resources through value restrictions in a subclass axiom.

The approach presented here allows for many to many mappings between multiple vocabularies and meets the requirements of section 4. It uses many of the new features in OWL 2: property chains, disjoint properties, and negative property assertions. The use of OWL 2 constructs provides a lot of freedom in specifying class restrictions beyond those presented here. For instance, the applicability of some mapping may depend on the applicability of another by a simple subclass relation. As presented here, the approach does not fit any of the OWL 2 profiles, but it is worth investigating whether the approach can be meaningfully supported by one of them [3]. With a mapping ontology in place, we can start the definition of mappings between layman concepts and legal tort concepts. It would be interesting to see whether the assumption that such mappings can improve accessibility holds, and especially what the effect of an additional step is on precision and recall.

Acknowledgements

The BEST project is funded under grant number 634.000.436 by the Netherlands Organisation for Scientific Research (NWO). See `http://www.best-project.nl` for more information.

References

1. McCulloch, E., Macgregor, G.: Analysis of equivalence mapping for terminology services. Journal of Information Science **34**(1) (2008) 70–92
2. Miles, A., Bechhofer, S.: SKOS simple knowledge organization system reference. Technical report, W3C (2009)

3. W3C OWL Working Group: OWL 2 web ontology language document overview. Technical report, World Wide Web Consortium (11th June 2009) Working Draft.
4. Isaac, A., Schlobach, S., Matthezing, H., Zinn, C.: Integrated access to cultural heritage resources through representation and alignment of controlled vocabularies. Library Review **57**(3) (2007) 187–199
5. Hoekstra, R.: BestPortal: Lessons learned in lightweight semantic access to court proceedings. In Governatori, G., ed.: Proceedings of the 22nd International Conference on Legal Knowledge and Information Systems (JURIX 2009). Frontiers of Artificial Intelligence, Rotterdam, Jurix, IOS Press (December 2009)
6. Uijttenbroek, E.M., Lodder, A.R., Klein, M.C., Wildeboer, G.R., van Steenbergen, W., Sie, R.L., Huygen, P.E., van Harmelen, F.: Retrieval of case law to provide layman with information about liability: Preliminary results of the best-project. In: Computable Models of the Law. Volume 4884/2008 of LNCS. (2008)
7. Winkels, R., de Bruijn, H.: Case frames: Bridging the gap between a case and the law. In Carr, I., Narayanan, A., eds.: Proceedings of the First European Conference on Computers, Law and AI, Exeter, EUCLID (1996) 205–213
8. Valente, A.: Legal Knowledge Engineering: A Modelling Approach. PhD thesis, University of Amsterdam, Amsterdam (1995)
9. Hoekstra, R., Breuker, J., Di Bello, M., Boer, A.: LKIF Core: Principled ontology development for the legal domain. In Breuker, J., Casanovas, P., Klein, M., Francesconi, E., eds.: Law, Ontologies and the Semantic Web. Frontiers of Artificial Intelligence and Applications. IOS Press, Amsterdam (2008)
10. Hoekstra, R.: Ontology Representation – Design Patterns and Ontologies that Make Sense. Volume 197 of Frontiers of Artificial Intelligence and Applications. IOS Press, Amsterdam (June 2009)
11. van de Ven, S., Hoekstra, R., Breuker, J., Wortel, L., El-Ali, A.: Judging Amy: Automated legal assessment using OWL 2. In: Proceedings of OWL: Experiences and Directions (OWLED 2008 EU). (October 2008)
12. van Assem, M., Malaisé, V., Miles, A., Schreiber, G.: A method to convert thesauri to skos. In Sure, Y., Domingue, J., eds.: ESWC. Volume 4011 of Lecture Notes in Computer Science., Springer (2006) 95–109
13. Isaac, A., van der Meij, L., Schlobach, S., Wang, S.: An empirical study of instance-based ontology matching. In: Proceedings of the 6th International Semantic Web Conference (ISWC 2007), Busan, Korea (2007)
14. Polleres, A., Scharffe, F., Schindlauer, R.: SPARQL++ for mapping between RDF vocabularies. In: OTM Confederated International Conferences CoopIS, DOA, ODBASE, GADA, and IS 2007. Volume 4803 of Lecture notes in computer science. (2007) 878
15. Prud'hommeaux, E., Seaborne, A.: SPARQL query language for RDF. Technical report, World Wide Web Consortium (2008)
16. Isaac, A., Wang, S., Zinn, C., Matthezing, H., van der Meij, L., Schlobach, S.: Evaluating thesaurus alignments for semantic interoperability in the library domain. IEEE Intelligent Systems **24**(2) (2009) 76–86
17. Motik, B., Patel-Schneider, P., Parsia, B., Bock, C., Fokoue, A., Haase, P., Hoekstra, R., Horrocks, I., Ruttenberg, A., Sattler, U., Smith, M.: OWL 2 web ontology language structural specification and functional-style syntax. Technical report, World Wide Web Consortium (11th June 2009) Candidate Recommendation.