Enhancing Linguistic Web Service Description with Non-functional NLP Properties

Nabil Baklouti[1], Bilel Gargouri[2] and Mohamed Jmaiel[1]
[1]*ReDCAD Laboratory, University of Sfax, Sfax, Tunisia*
[2]*MIRACL Laboratory, University of Sfax, Sfax, Tunisia*

Keywords: Linguistic Web Service, Non-functional NLP Properties, Semantic Web, Ontology, OWL-S.

Abstract: This paper deals with the enhancing of Linguistic Web Service (LingWS) description. It proposes an extension for the OWL-S approach and a Natural Language Processing (NLP) domain ontology based on linguistic standards. The proposed extension provides a classification of the Non-functional NLP properties which promotes the representation of their relationships. The extended OWL-S description is linked to the NLP domain ontology to semantically annotate the LingWS properties.

1 INTRODUCTION

The lingware system engineering is a sub-domain of the Software Engineering related to Natural Language Processing (NLP). The development of this kind of systems needs both several linguistic resources and treatments. For this, researchers in NLP have resorted to reuse existing lingware systems. These attempts are based on the Web Service technology such as (Ishida, 2006), (Tufis et al., 2008), and (Baklouti et al., 2010).

Today, many Linguistic Web Services (LingWS) are published on the internet. They deal with various applications such as Question/Response system and Information Retrieval (Bramantoro et al., 2008). The LingWS can be used like a simple Web Service or integrated to a composite Web Service.

The Web Service discovery is the process of locating one or more related documents that describe a particular Web service using the Web Services Description Language (WSDL) (W3C, 2001) which represents a Service-Oriented Architecture (SOA) standard. Moreover, the SOA uses SOAP (Simple Object Access Protocol) for the invocation and UDDI (Universal Description Discovery and Integration) for publishing Web Services. Unfortunately, these standards do not provide efficient discovery results since they are essentially focused on syntactical description of Web Service. For example, the UDDI registry organizes the services according to defined categories without introducing semantic aspects. UDDI's search is also based on syntax like service's name and it relies on XML. The lack of semantics inside SOA standards powerfully influences the LingWS discovery. In fact, there is a lack of NLP properties and their relations. For example, we should know which phenomena are covered by a LingWS and what related approaches and resources are used. According to the literature, several works suggested adding a semantic wrapper or an NLP domain ontology for improving the description of LingWS. However, the Non-Functional NLP properties and their relationships are not considered in the above works.

Hayashi (Hayashi, 2011) has asserted that researchers in the NLP field have to develop an external mechanism to semantically enrich the LingWS descriptions in order to enhance its discovery.

The LingWS description should cover the Non-Functional NLP properties and relationships between them like the performed treatment type, the resource, the analysis type, and so on. Unfortunately, the existing semantic approaches such as OWL-S (Martin et al., 2004), WSMO (essi WSMO working group, 2004), and SAWSDL (Farrell and Lausen, 2007) are unable to represent this kind of properties and their relations.

In other domains, some extensions of OWL-S approach have been proposed such as (Aier et al., 2007) and (Jean et al., 2010). These extensions aim to integrate the quality standards of Web Services.

This paper proposes a semantic enrichment of LingWS description by integrating the Non-Functional NLP properties and their relations. For this, we make an extension for OWL-S approach

Baklouti N., Gargouri B. and Jmaiel M..
Enhancing Linguistic Web Service Description with Non-functional NLP Properties.
DOI: 10.5220/0004080604390444
In *Proceedings of the 7th International Conference on Software Paradigm Trends* (ICSOFT-2012), pages 439-444
ISBN: 978-989-8565-19-8

which aims to both classify Non-functional NLP properties and represent their relationships. In addition, we build a domain ontology using the NLP norms to annotate the LingWS description. Thus, the proposed solution is distinguished by allowing the description of the NLP specificities in a separate way while highlighting linguistic links between them.

The remaining of this paper is structured as follows. Section 2 shows the attempts related to the LingWS discovery problems. In section 3, we present the Non-functional NLP properties. The proposed extension is given in section 4. Section 5 shows the NLP domain ontology. In section 6, we provide a practical study to illustrate how the proposed extension can represent the Non-functional NLP properties and their relations. The last section concludes the paper.

2 RELATED WORK

We divide this section into two parts. We start with presenting the relevant works in the NLP related to the LingWS discovery. In the second part, we make a comparative study of the semantic approaches.

In order to enhance the discovery of LingWS, there are some relevant works which have proposed to associate a wrapper around LingWS using semantic technologies (i.e., OWL[1] and OWL-S) such as (Ishida, 2006). It represents the LingWS Profile which contains the LingWS Name, the LingWS Type, a textual description, LingWS Status, and so on. However, this profile does not contain other relevant Non-functional NLP properties and mainly their relations which may improve the LingWS discovery. Another issue is the absence of an ontology which represents both linguisic processing resources and their Input/Output (I/O).

Klein and Potter (Klein and Potter, 2004) have proposed an ontology for describing LingWS using OWL-S approach. For our knowledge, it had explicitly stated the necessity of ontological foundation for language infrastructure. Nevertheless, this proposition ignores taxonomies for both language resources and abstract objects. In addition, the OWL-S is unable to both classify Non-functional NLP properties and establish relationships between them. In terms of linguistic resources interoperability, this proposition does not take into account any NLP standard (e.g., LMF).

Hayashi (Hayashi, 2011) proposed an ”ontologization” of the Lexical Markup Framework (LMF[2]). This work does not represent the LingWS I/O which are important for ensuring LingWS discovery. Hayashi et al. (Hayashi et al., 2008),(Hayashi, 2011) used SAWSDL to annotate LingWS description. Nevertheless, SAWSDL cannot represent the details of the NLP knowledge. Moreover, Hayashi in (Hayashi, 2011) has asserted that researchers in the NLP field have to develop a mechanism for discovering allowing the semantic enrichment of LingWS description.

To conclude, the LingWS description should be augmented with Non-functional properties and their relationships which can enhance the discovery task. In addition, the NLP domain ontology should be more expressive in terms of NLP specificities.

In order to overcome the WSDL semantic lack, various approaches have been proposed such as OWL-S (Martin et al., 2004), WSMO (essi WSMO working group, 2004), and SAWSDL (Farrell and Lausen, 2007).

The OWL-S approach is built inside the Web Services. It proposes an ontology of services motivated by the need to provide three elements: The **Profile** which is used to announce the service. It contains the I/O, the preconditions, the results, and the service category of Web Service. The **Process** which contains I/O, preconditions, results, and the behaviour of the service (data and control flow), and the third element is the **Grounding** which provides the details (e.g, protocol, address) to invoke the services.

The Web Service Modelling Ontology (WSMO) provides a framework for semantic descriptions of Web Services and acts as a meta-model for such Services based on the Meta Object Facility (MOF) [3]. Semantic service descriptions, according to the WSMO model, can be defined using Web Service Modeling Language (WSML[4]). It consists of four elements deemed necessary to support Semantic Web services: **Ontologies**, **Goals**, **Web Services**, and **Mediators**.

Semantic Annotations for WSDL and XML Schema (SAWSDL) defines a new name-space called ”sawsdl”. There are three extensions for it : **modelreference** which associates an XML Schema or a WSDL component to an ontology concept. The other two extensions are **liftingSchemaMapping** and **loweringSchemaMapping** which promote the mapping between the semantic data and the XML elements.

We use some criteria to compare the above approaches: For SAWSDL, it deals with the discovery and the automatic invocation of Web Services but not for the composition. With SAWSDL, we can use any type of ontology (e.g, OWL, WSML) while OWL-S supports only OWL ontologies and WSMO

[1] http://www.w3.org/2004/OWL/

[2] http://www.lexicalmarkupframework.org/

[3] http://www.omg.org/mof/

[4] http://www.wsmo.org/TR/d16/d16.1/v0.21/

the WSML ontologies. Otherwise, the SAWSDL does not promote the definition of Non-Functional properties. In terms of tools, OWL-S provides several tools such as editor, matcher, and composer. The WSMO tools are more difficult to develop because they are based on WSML which is never used in the past nevertheless OWL-S and SAWSDL rely on RDF and XML.

We clearly note that OWL-S approach provides more benefits in terms of both ontology, tools, Non-Functional properties, and composition. Thereafter, OWL-S seems to be the best approach for describing LingWS.

3 NON-FUNCTIONAL NLP PROPERTIES

Several types of information need to be modelled within the LingWS description. For this, we present in the following some Non-functional NLP properties. We classify these properties by processing level:

- **Lexical Level:** It is characterized by the use of lexical linguistic resources, lexical approaches, lexical formalisms, lexical analysis types and so on.
  For example, the developer can choose its LingWS according to the used lexical analysis such as the thematic (It proposes a large conceptual category in which the user can navigate for finding the suitable word), the structural (It helps the writer in the structure choice) or the syntagmatic one (It is a statement element regrouped into sub-parties with an internal structure and a coherent unit).
- **Morphological Level:** It contains various features such as morphological phenomena, morphological formalisms, and approaches. As an example, there are the Linguistic approach (It segments a text to elementary units which have a linguistic knowledge attached: grammatical category , gender, number , time, person and so on.), the Statistic approach (The analysis starts by splitting sentences into words. Then, a cost is attributed to each bi-gramme according to the calculated aparition frequency in a corpus. Finally, the solution which has the lower cost is chosen like the best probable.), and the Hybrid approach (It combines linguistic and statistic criteria. It extracts the relevant terms from both text statistic analysis and linguistic filtering of the candidate terms. It produces a sorted list of the most representative terms for a specific domain.).
  We show the requirement of the approach kind by the following example : When a developer has the intention to build a morphological application which uses a Linguistic approach, so he has to take a way all LingWS using other kinds of approaches.
- **Syntactic Level:** Different specificities can characterize a syntactic LingWS such as syntactic phenomena, syntactic analysis, and syntactic formalisms. As an example, for the analysis type we can mention: Top-down analysis (The analysis begins from the start symbol called axiom and try to rebuild the derivation tree by a prefixed left-right course.), Bottom-Up analysis (It factorizes the word by picking out or recognizing the right parts of production until find out the axiom.), Profound analysis (It produces a formal representation of the sentences, under a syntactic tree form.), Surface or Chunking (It identifies the components limits i.e., Nominal Group (NG) and Verbal Group (VG)) and Structural analysis (It is based on a set of rules for defining associations between words in order to construct sentences.).
  When an application treats two phenomena (e.g., Accord and Anaphora) using two formalisms (e.g., Unification Grammar and Resolution Algorithm), so the developer has to choose the suitable LingWS according to the used formalism for each retained phenomenon.
- **Semantic Level:** To develop a semantic application, we can choose some NLP properties such as how to represent knowledge, the semantic formalisms, the semantic phenomena, and the semantic resources. The used resource is a relevant information. Indeed, if a developer wants to compose an application which needs a Wordnet resource then he has to eliminate LingWS using LMF resource for example.

After presenting the Non-Functional NLP properties, we observe that a LingWS description should contain this kind of properties for enhancing LingWS discovery.

4 EXTENSION OF OWL-S

4.1 Drawbacks of OWL-S

The **Profile** promotes the description of both Functional and Non-Functional properties of Web Services. It is associated to a set of **ServiceParameter** which allows the annotation of services by couples (criteria, value). However, the latter cannot be used to

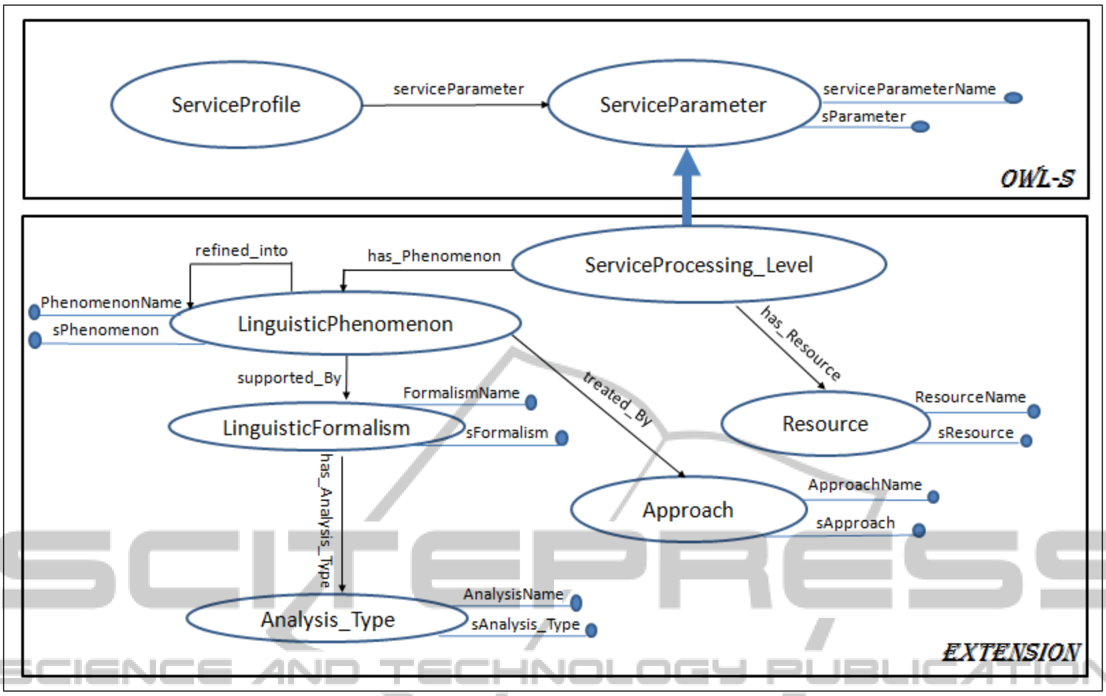

Figure 1: Extension of OWL-S for integrating Non-functional NLP properties.

represent relationships between Non-functional NLP properties. In fact, it is not possible to: (1) Use such criterion to classify the Non-functional NLP properties, so using OWL-S model and for one LingWS, we cannot obtain both processing level and its linguistic properties such as phenomena, resources, and formalisms. (2) Remove ambiguity: If a LingWS covers many phenomena and each one is treated by one formalism, so we cannot keep these relations. (1) and (2) prove that an extension of OWL-S can be very useful to further exploit the specified Non-functional NLP properties of LingWS.

4.2 Proposed Extension

The OWL-S extension is presented by Figure 1. As it is shown by this figure, the OWL-S extension is based on the specialization of 'ServiceParameter' class by one class namely 'ServiceProcessing_Level'.
The main elements of the proposed extension are:

- *ServiceProcessing_Level*: It represents the processing level of the LingWS. We have essentially four processing levels which are Syntactic, Lexical, Morphological, and Semantic. Each processing level is characterized by both its resources (e.g., dictionaries, tree bank, corpus) and phenomena (e.g., ellipsis, anaphora, accord). For this reason, we add respectively the *Resource* and the *LinguisticPhenomenon* classes.
- *LinguisticPhenomenon*: It has the 'refined_into' relation, since each phenomenon has its sub-Phenomena. For example, for the ellipsis phenomenon, we can find the nominal ellipsis (the omission of the essential part of a nominal phrase: the head) and an ellipsis of a whole phrase (e.g., subject ellipsis, verb ellipsis, both verb and complement ellipsis). The *LinguisticPhenomenon* has also other relations with other classes:
  - *Approach*: the treatment approach of a such phenomenon.
  - *LinguisticFormalism*: represents the formalism (e.g., HPSG and LFG for syntactic Grammars). It supports the phenomenon determination using an *Analysis_Type* (e.g., Top-Down, Bottom-Up).

5 NLP DOMAIN ONTOLOGY

It is essential to develop an NLP domain ontology to annotate the elements of the LingWS description (e.g, I/O and the recently added properties). Thus, the proposed extension and the NLP ontology should be linked using pointers (e.g, sParameter) in order to give an expressive semantic description. Unfortunately, the existing ontologies are incomplete and not specialized to improve the LingWS discovery (Hayashi

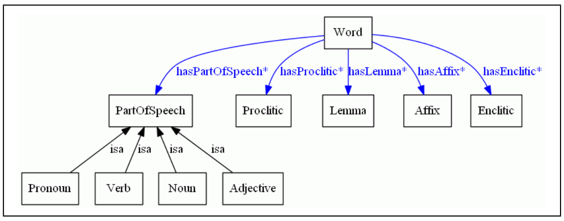

Figure 2: Extract from NLP domain ontology.

et al., 2008). Hence, we have developed a new ontology which both promotes linguistic resource taxonomies and takes advantages of ISO standards proposed in the area of lexical resources construction, namely LMF (ISO 24613) and DCR (ISO 12620). We have used these standards to identify concepts and data-properties (i.e, relations). The developed ontology is extensible, so we can add other resources (i.e., data resource and processing resource). We used OWL[5] as description language and Protégé[6] as the main tool for the ontology construction. Figure 2 shows an extract of the developed ontology dealing with the Morphological and the Syntactic levels.

6 DEMONSTRATION

To consolidate our solution, we use a service library available in our laboratory which contains many LingWS then, we take one for applying the OWL-S extension using the NLP domain ontology. Currently, our library contains about 40 LingWS which can be expanded by other. The available LingWS cover some languages : Arabic, French and English. We obtained these LingWS from both (Baklouti et al., 2010) and from some opensource tools like OPEN-NLP[7], NLP-LIB, classifier4j, standford[8], and so on. The major of these tools are used by the known platforms (e.g, GATE and UIMA).

We choose the 'Syntactic_Parser' as an example of LingWS for making a practical study. The latter promotes many tasks such as it performs context-free syntax analysis, it guides context-sensitive analysis, and it attempts error correction.

Now, we present the 'Syntactic_Parser' profile using the proposed extension.

For this, we use OWL-S Editor[9] which is a plugin for protégé. It is the most used tool for OWL-S, since it performs several tasks such as edition, composition, and execution. Moreover, it is able to both load WSDL files of existing LingWS, generate an OWL-S 'skeleton' using wsdl2owls class, annotate generated description, and execute Semantic LingWS.

Figure 3 shows how we represent the Non-Functional NLP properties as well as the relationships between them. This representation is not possible using the initial OWL-S profile. To do this, we added some 'Object Properties' like 'refined_into', 'treated_By', and so on.

As it is indicated by Figure 3, the 'Syntactic_Parser' LingWS treats 'Ellipsis' as 'LinguisticPhenomenon' which is refined into two subphenomena: 'Nominal_Ellipsis' and 'Verb_Ellipsis'. The first one is resolved using 'Semantic_Approach' and the second one by an 'Hybrid_Approach' as 'Approach'.

```
<profile:LinguisticPhenomenon rdf:ID="LinguisticPhenomenon_1">
 <profile:PhenomenonName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Ellipsis
</profile:PhenomenonName>
 <sPhenomenon rdf:resource="http://www.owl-ontologies.com/Ontology1326481497.owl#Phenomenon"/>
  <profile:refined_into>
   <profile:LinguisticPhenomenon rdf:ID="LinguisticPhenomenon_2">
   <profile:PhenomenonName rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Nominal_Ellipsis</profile:PhenomenonName>
    <sPhenomenon rdf:resource="http://www.owl-ontologies.com/Ontology1326481497.owl#Phenomenon"/>
    <profile:treated_By>
     <profile:Approach rdf:ID="Approach_1">
      <sApproach rdf:resource="http://www.owl-ontologies.com/Ontology1326481497.owl#Approach"/>
      <profile:ApproachName rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Semantic_Approach</profile:ApproachName>
     </profile:Approach>
    </profile:treated_By>
    <profile:LinguisticPhenomenon rdf:ID="LinguisticPhenomenon_3">
     <profile:PhenomenonName
         rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Verb_Ellipsis</profile:PhenomenonName>
         <sPhenomenon rdf:resource="http://www.owl-ontologies.com/Ontology1326481497.owl#Phenomenon"/>
         <profile:treated_By>
           <profile:Approach rdf:ID="Approach_2">
                  <sApproach rdf:resource="http://www.owl-ontologies.com/Ontology1326481497.owl#Approach"/>
                  <profile:ApproachName rdf:datatype= "http://www.w3.org/2001/XMLSchema#string">
                  Hybrid_Approach</profile:ApproachName>
           </profile:Approach>
```

Figure 3: OWL-S extension code of 'Syntactic_Parser'

Figure 3 also shows the links between the Non-Functional NLP properties and the NLP domain ontology by respectively 'sPhenomenon' for 'LinguisticPhenomenon' and 'sApproach' for 'Approach'.

This figure does not contain the initial OWL-S elements such as 'serviceName', 'textDescription' but only the recently added NLP properties (i.e, 'LinguisticPhenomenon', 'Approach').

This section clearly prove how the proposed extension can both represent the Non-functional NLP properties and establish relationships between them.

[5]http://www.w3.org/TR/2009/REC-owl2-overview-20091027/

[6]http://protege.stanford.edu/

[7]http://incubator.apache.org/opennlp/

[8]http://nlp.stanford.edu/links/statnlp.html

[9]http://owlseditor.semwebcentral.org/index.shtml

7 CONCLUSIONS AND FUTURE WORK

We provided a solution to the problems related to the lack of semantic within the LingWS description. Indeed, we proposed an OWL-S extension for integrating Non-functional NLP properties and relations between them, since they can semantically enhance the LingWS descriptions. Besides, we developed an NLP domain ontology using NLP ISO standards.

The demonstration that we made proves the effectiveness of the OWL-S extension. Actually, the LingWS description is more meaningful, so it promotes advanced research (i.e, many combinations) knowing also that each property can be accessible alone.

For the future, we plan to cover other NLP properties mainly for the non-Latin languages. Then, we will check out the LingWS discovery by defining an appropriate matching algorithm. Finally, we will address the LingWS composition.

ACKNOWLEDGEMENTS

A Part of this work was done during a SIMBAD supported internship at the TELECOM SudParis, Evry France. For this, the authors thank Samir Tata for its support.

REFERENCES

Aier, S., Offermann, P., Schonherr, M., and Schropfer, C. (2007). Implementing non-functional service descriptions in soas. In *Trends in Enterprise Application Architecture Lecture Notes in Computer Science*, volume 4473/2007, pages 40–53.

Baklouti, N., Bouaziz, S., Gargouri, B., Aloulou, C., and Jmaiel, M. (2010). Towards the reuse of lingware systems: A proposed approach with a practical experiment. In *Proceedings of the International Conference on Information Integration and Web-based Application and Services (iiWAS2010) ACM*, pages 566–572, Paris-France.

Bramantoro, A., Tanaka, M., Murakami, Y., Schafer, U., and Ishida, T. (2008). A hybrid integrated architecture for language service composition. In *IEEE International Conference on Web Services*.

essi WSMO working group (2004). Web service modeling ontology. http://www.wsmo.org/. http://www.wsmo.org/.

Farrell, J. and Lausen, H. (2007). Semantic annotations for wsdl and xml schema. http://www.w3.org/TR/sawsdl. http://www.w3.org/TR/sawsdl.

Hayashi, Y. (2011). Prospects for an ontology-grounded language service infrastructure. In *Proceedings of Workshop on Language Resources, Technology and Services in the Sharing Paradigm*, page 17, Chiang Mai, Thailand.

Hayashi, Y., Narawa, C., Monachini, M., Soria, C., and Calzolari, N. (2008). Ontologizing lexicon access functions based on an lmf-based lexicon taxonomy. In *LREC*, Marrakech, Morocco. European Language Resources Association (ELRA).

Ishida, T. (2006). Language grid: An infrastructure for intercultural collaboration. In *proceedings of the International Symposium on Applications and the Internet (SAINT06) IEEE*, pages 96–100.

Jean, S., Losavio, F., Matteo, A., and Lévy, N. (2010). An extension of owl-s with quality standards. In *Proceedings of the Fourth IEEE International Conference on Research Challenges in Information Science*, pages 483–494.

Klein, E. and Potter, S. (2004). An ontology for nlp services. In *Proc. of LREC Workshop on a Registry of Linguistic Data Categories within an Integrated Language Resource Repository Area*.

Martin, D., Paolucci, M., McIlraith, S., Burstein, M., McDermott, D., McGuinness, D., Parsia, B., Payne, T., Sabou, M., Solanki, M., Srinivasan, N., and Sycara, K. (2004). Bringing semantics to web services: The owl-s approach. In *International Workshop on Semantic Web Services and Web Process Composition*.

Tufis, D., Ion, R., Ceausu, A., and Stefnescu, D. (2008). Racais linguistic web services. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC08)*, Marrakech, Morocco. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2008/.

W3C (2001). Web services description language (wsdl). http://www.w3.org/TR/wsdl/. http://www.w3.org/TR/wsdl/.