Describing bibliographic references in RDF

Angelo Di Iorio[1], Andrea Giovanni Nuzzolese[1,2], Silvio Peroni[1,2], David Shotton[3], and Fabio Vitali[1]

[1] Department of Computer Science and Engineering, University of Bologna (Italy)
[2] STLab-ISTC, Consiglio Nazionale delle Ricerche (Italy)
[3] Oxford e-Research Centre, University of Oxford (UK)
angelo.diiorio@unibo.it, nuzzoles@cs.unibo.it, silvio.peroni@unibo.it, david.shotton@oerc.ox.ac.uk, fabio@cs.unibo.it

Abstract. In this paper we present two ontologies, i.e., *BiRO* and *C4O*, that allow users to describe bibliographic references in an accurate way, and we introduce *REnhancer*, a proof-of-concept implementation of a converter that takes as input a raw-text list of references and produces an RDF dataset according to the *BiRO* and *C4O* ontologies.

Keywords: BiRO, C4O, REnhancer, SPAR, Semantic Publishing, bibliographic references, citation network

1 Introduction

Within the scholarly domain, the Semantic Publishing is the use of Semantic Web technologies to enhance published scholarly articles. In the seminal paper [17], Shotton suggests that the road to the Semantic Publishing will proceed through incremental steps, starting simply and then improving the whole functionality of the system. The (semantic) management of bibliographic references is one of the areas where such an incremental approach is most feasible and effective.

In [15] we presented a "manifesto" on liberating bibliographic references. The bibliographic references are core elements of scholarly communication – since they permit the attribution of credit and integrate our independent research endeavours – and must be freely available for use by scholars. Citation data "should be recognised as a part of the Commons" [15] – i.e., data that are freely and legally available for sharing – and placed in an open repository, stored in appropriate machine-readable formats, so as to be easily reused by machines to assist people in producing novel services.

Creating such open repositories of bibliographic references is not an easy task. First of all, one should create an appropriate (and unique) description of each referenced document, probably starting from several bibliographic references that are written according to different schemas[4]. Also, bibliographic references are difficult to normalise as they could contain typos on all elements of the reference (article title, authors' information, etc.). The format these references are

[4] A large repository of over 6000 styles for references is available (in CSL 1.0 format) at https://www.zotero.org/styles.

exported is also an issue. The bibliographic entries can already be exported as textual content or as record-like structures, but only a few repositories export their data using Semantic Web technologies, such as Nature with its Linked Data Platform[5] and the Open Citation Corpus[6] [18].

Thus, ontologies to model references and reference lists are needed. It is crucial that these ontologies cover all traits of bibliographic references and handle them appropriately. For instance, a critical point is the distinction between the references in the reference lists and the actual cited articles, as well as the distinction between the in-line pointers to the reference and the actual reference (which, in turn, is different from the cited article).

In this paper we present two ontologies – i.e., the *Bibliographic Reference Ontology* (*BiRO*) and the *Context Characterisation and Citation Counting Ontology* (*C4O*) – addressing these issues. They are two of the Semantic Publishing And Referencing (SPAR) Ontologies[7], a suite of ontologies that describe the different aspects of the scholarly publishing domain. In particular, BiRO and C4O have been developed for describing bibliographic lists, bibliographic references, in-text reference pointers, citation contexts and a mechanism for counting citations locally (within an article) or globally (by means of particular platforms).

The definition of ontologies for bibliographic resources is not enough, if it is not combined with tools for populating these ontologies (cf. [15]). In order for more actors to fully participate to the Semantic Publishing evolution and, in particular, to allow authors and readers to give a strong contribution ("[publishers,] expect greater things of your author" [17]), we implemented a proof-of-concept tool that helps users to convert textual lists of references into RDF descriptions, compliant with BiRO and C4O. The prototype is named REnhancer and is briefly presented at the end of the paper.

The rest of the paper is structured as follows. In Section 2 we provide some clarifications about the nomenclature related to bibliographic references, as well as some related works on this topic. In Section 3 and Section 4 we introduce BiRO and C4O respectively, and we show how they can be used to describe bibliographic references and related objects. In Section 5 we introduce REnhancer and how to use its Web interface and its REST API. We conclude the paper in Section 6 sketching out our future works on this topic.

2 What is a citation really?

The word "citation" is often subject of misinterpretations and misuse. The reason being that the word can be used to identify objects which have different purposes, at least in scientific literature. For instance, we often identify as "citation" (a) the *act of citing* another work, (b) a *bibliographic reference* put at the end of a paper (usually in a list), as well as (c) particular *pointers* (e.g., "[3]") denoting that bibliographic reference.

[5] Nature.com Linked Data: http://data.nature.com.
[6] Open Citation Corpus: http://opencitations.net.
[7] Semantic Publishing And Referencing (SPAR) Ontologies: http://purl.org/spar.

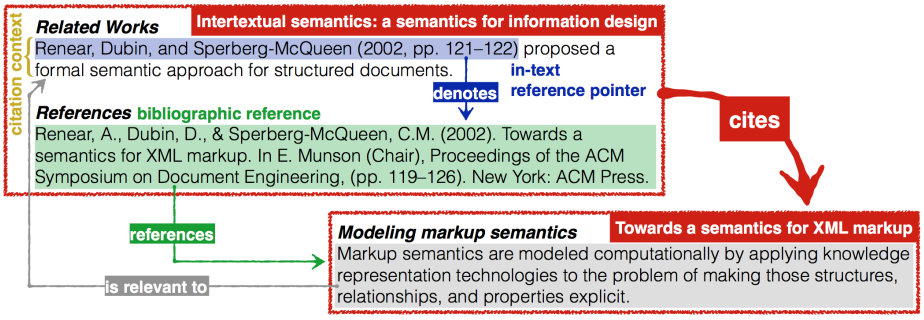

Fig. 1. An excerpt of [12] highlighting the different roles of the various parts of text.

In order to expand more on this topic, let us consider the excerpt from the article "Intertextual semantics: a semantics for information design" [12] shown in Fig. 1. That excerpt contains a particular sentence from the section "Related Works" of the paper and a list item from the final "References" section. In [15], we identified three different kinds of objects that are used to express any citation (i.e., the attribution link between a citing work and the cited work) that are relevant for this work, all of them having different purposes:

- *bibliographic reference*, i.e., the textual entity within a citing work that references a cited work. A bibliographic reference is the realisation of some bibliographic record for the cited work, arranged in a specific format determined by the house style of the citing publication;
- *in-text reference pointer*, i.e., the entity present in the body text of a citing work that denotes a particular bibliographic reference in the reference list or a footnote. In scientific literature, this in-text reference pointer can be presented in different forms;
- *citation context*, i.e., the textual content of that component of the published paper (e.g., sentence, paragraph) within which an in-text reference pointer appears, which provides the rhetorical rationale for the existence of that citation. In can includes the sentence where it appears (i.e., *citation sentence* [2]), a sequence of consecutive sentences referring (explicitly or implicitly) to the same cited work (i.e., *context window* [16]), and the main structure containing the in-text reference pointer (e.g., the paragraph).

However, to our knowledge, there are no ontologies that have been developed to model all these objects, since most of them usually focus on describing the whole citing/cited entity. In the following we will briefly describe the most relevant ontologies in this context and their capabilities.

Dublin Core Metadata Terms (DCTerms) [5] is, as far as we know, the most widely used vocabulary for describing and cataloguing resources. While very useful for the creation of basic metadata for resource discovery, the main limitation of DCTerms is a direct consequence of the generic nature of its terms. For example, using DCTerms one can identify a creator but not an author; a

bibliographic resource but not a journal article; an identifier but not an ISSN, and a date but not a publication date. However, it makes available a particular property, i.e., *dcterms:bibliographicCitation*, to define the textual string describing a reference, and another property, i.e., *dcterms:references*, to indicate that an entity A cites/points another entity B.

Similar to DCTerms, the RDF specification of the *Publishing Requirements for Industry Standard Metadata* (*PRISM*) [10] has an extensive set of terms for the description of bibliographic entities that is richer than DCTerms (its main limitation is that it is a flat structure**,** lacking hierarchies). It makes available the property *prism:references* (and its inverse *prism:isReferencedBy*) to define citations between entities.

The *Bibliographic Ontology* (*BIBO*) [8], is an OWL Full ontology that allows one to write descriptions of documents (*bibo:Document* is the core class of that model) for publication on the Semantic Web. It includes both DCTerms and PRISM properties to cover common needs, and it adds other classes and properties to describe in more detail the publishing domain. In particular, it explicitly defines the property *bibo:cites* to express citations between *documents*.

Among the SPAR ontologies, FaBiO, CiTO [13] and DoCO [7] are ontologies that make available a first infrastructure to organise the citation network between scholarly articles. In particular, FaBiO has a class, i.e., *fabio:Bibliographic Metadata*, that enable the description of usual metadata associated to scholarly articles (e.g., authors' names, title, journal name, DOI); CiTO is basically a list of properties defining citation acts between generic entities (e.g., *cito:extends*, *cito:uses MethodIn*, *cito:disagreesWith*); and, finally, DoCO makes available classes defining document components for the characterisation of bibliographic references, i.e., *deo:BibliographicReference* and *doco:BibliographicReferenceList*.

In conclusion, none of these ontologies provides all entities (classes and properties) useful to prevent or minimise ambiguities when modelling citing acts in documents. In the next sections we will go into details of two ontologies that can be combined to overcome these limitations: BiRO and C4O.

3 Describing bibliographic references: BiRO

The *Bibliographic Reference Ontology*[8] (*BiRO*) describes reference lists and references by using Semantic Web technologies. In particular, BiRO uses an OWL-based definition of the FRBR model (prefix *frbr*)[9] to define bibliographic references and their compilation into ordered bibliographic lists, by means of the *Collections Ontology* (prefix *co*)[10] [3], as shown in Fig. 2.

An individual bibliographic reference, such as one in the reference list of a published journal article, may exhibit varying degrees of incompleteness, depending on the formatting rules of the journal. For example, it may lack the title

[8] BiRO, the Bibliographic Reference Ontology: http://purl.org/spar/biro.
[9] FRBR ontology: http://purl.org/spar/frbr.
[10] CO, the Collections Ontology: http://purl.org/co.

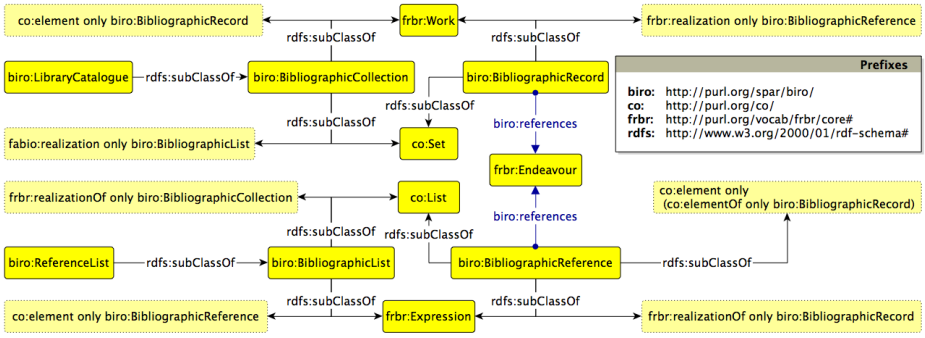

Fig. 2. Graffoo diagram (http://www.essepuntato.it/graffoo) summarising the *Bibliographic Reference Ontology* (*BiRO*).

of the cited article, the full names of the listed authors, or indeed a full listing of the authors.

BiRO provides a logical system for relating such incomplete bibliographic reference to:

- the full bibliographic record for that cited article, which, in addition to any author and title fields missing from the reference, may also be expected to include the name of the publisher, and the ISSN or ISBN of the publication;
- collections of bibliographic records, such as library catalogues; and
- ordered bibliographic lists, such as reference lists.

In order to understand how to use BiRO to describe reference lists, let us take into account again the reference introduced in Fig. 1 referring to Renear *et al.*'s paper.

A first way for defining a simple machine-readable representation of that reference using BiRO with FRBR and DCTerms is as follows:

```
:intertextual-semantics frbr:part :reference-list .
:reference-list a biro:ReferenceList ;
  co:firstItem [ co:itemContent :barwise83 ;
  co:nextItem [ co:itemContent :black37 ; ...
  co:nextItem [ co:itemContent :renear02 ; ... ] ... ] ] . ...
:renear02 a biro:BibliographicReference ;
  dcterms:bibliographicCitation "Renear, A., Dubin, D. & Sperberg-McQueen, C.
       M. (2002). Towards a semantics for XML markup. In E. Mudson (Chair),
       Proceedings of the ACM Symposium on Document Engineering, (pp. 119-126)
       . New York: ACM Press." . ...
```

This formal description is not fully expressive as we only assigned an IRI to the reference list and to each of its references, and the semantics of the string representing the reference is still obscure. For instance, there is no explicit statement saying that the strings "Renear, A.", "2002" and "Towards a semantics for XML markup" are, respectively, the name of one of the authors, the year of publication and the title of the article. On the other hand, this is a first necessary step to release bibliographic references in RDF. We also discuss below two main approaches to associate a meaning to the strings composing the reference.

Note also that if a complete RDF description of an article has already been created, even according to a different ontological model, we can use the property *biro:references* to create an explicit link between a reference citing that article and the article itself, or better its description. The following excerpt, for instance, shows how to say that the reference whose IRI is *:renear02* references the article whose IRI is *:towards-a-semantics*:

```
:renear02 biro:references :towards-a-semantics .
```

3.1 Semantic enhancement of strings: literal reification

A way to enable the semantic enhancement of strings, and to solve the above mentioned limitations, is to use literals as subjects of assertions, by promoting them as "first class object" in OWL. The pattern *literal reification* (prefix *literal*)[11] [9] fulfils this scenario by reifying literals as proper individuals of the class *litre:Literal*. Individuals of this class express literal values through the functional data property *litre:hasLiteralValue* and can be connected to other individuals that share the same literal value by using the property *litre:hasSameLiteralValueAs*. Moreover, a literal may refer to, and may be referred by, any OWL individual through *litre:isLiteralOf* and *litre:hasLiteral* respectively.

This pattern allows one to describe each string of a bibliographic reference as item of an ordered list of strings, by means of the Collections Ontology [3]. By means of this pattern and of the OWL 2 capabilities in meta-modelling, it becomes possible to link specific strings in the references and to enhance them through semantic assertions according to specific vocabularies, as shown in the following excerpt:

```
:renear02 a biro:BibliographicReference ;
  co:firstItem [ co:itemContent :first-author-name ; ...
  co:nextItem [ co:itemContent :publication-year ;
  co:nextItem [ co:itemContent :paper-title ; ... ] ] ] .
:first-author-name a literal:Literal , foaf:name ;
  literal:hasLiteralValue "Renear, A."^^xsd:string ;
  # it is the URL identifying the person referred by the above string
  literal:isLiteralOf :renear . ...
```

As shown above, now the bibliographic reference under consideration is described as a list of literals, each of them having a particular semantic connotation.

3.2 Semantic enhancement of strings: EARMARK ranges

Another approach to deal with the semantic enhancement of bibliographic references is to use EARMARK[12] ranges for associating appropriate semantic statements to textual fragments, as illustrated in [14]. For instance, let us encode the document cited in our example as an EARMARK document. We first need a particular string container (called *docuverse*) defining the text of the reference:

[11] Literal reification pattern: http://www.essepuntato.it/2010/06/literalreification. The prefix *literal* refers to entities defined in it.

[12] EARMARK ontology: http://www.essepuntato.it/2008/12/earmark. The prefix *earmark* refers to entities defined in it.

```
:renear02-reference a earmark:StringDocuverse ;
  earmark:hasContent "Renear, A., Dubin, D. & Sperberg-McQueen, C.M. (2002).
       Towards a semantics for XML markup. In E. Mudson (Chair), Proceedings
       of the ACM Symposium on Document Engineering, (pp. 119-126). New York:
       ACM Press." .
```

Then, we define ranges for each string we want to use in order to describe the bibliographic reference according to BiRO. These ranges can be defined as follows:

```
:renear02 a biro:BibliographicReference ;
  co:firstItem [ co:itemContent :first-author-name ;
  co:nextItem [ co:itemContent :publication-year ;
  co:nextItem [ co:itemContent :paper-title ; ... ] ] ] ] .
:first-author-name a earmark:PointerRange ; # the string "Renear, A."
  earmark:refersTo :renear02-reference ;
  earmark:begins "0"^^xsd:nonNegativeInteger ;
  earmark:ends "9"^^xsd:nonNegativeInteger . ...
```

Furthermore, using the *Linguistic Act* ontology [14], it is possible to link EARMARK ranges to their formal meaning and to the particular object referenced by such strings, as described in [1]. For instance, considering the range *:first-author-name*, we can say that:

- this range denotes a particular concrete object, i.e., a particular person identified by *:renear*;
- this range expresses a particular meaning, i.e., the fact that the string (as well as the denoted object) refers to something being an author of that paper.

Thus, we can express these additional assertions:

```
:first-author-name la:denotes :renear ; la:expresses
  [ a owl:Restriction ;
    owl:onProperty pro:holdsRoleInTime ;
    owl:someValuesFrom [ owl:intersectionOf (
      [ a owl:Restriction ; owl:onProperty pro:withRole ;
        owl:hasValue pro:author ]
      [ a owl:Restriction ; owl:onProperty pro:refersToDocument ;
        owl:hasValue :towards-a-semantics ] ) ] .
```

In this way we are able to identify in RDF the various part that form the reference and their specific meaning, and to link them to other entities.

4 What, where and how many times is cited: C4O

Besides defining reference lists and bibliographic references in a machine-readable form, it is also useful to describe how these references are used in the citing paper. In particular, we would need entities that describe:

- in-text reference pointers within the citing paper;
- links to the bibliographic references denoted by in-text reference pointers;
- how much a particular document is locally cited by the citing document – i.e., the total number of in-text reference pointers within the citing paper denoting the same bibliographic reference (that is useful for certain studies on citations [11]);

- how much an article is globally cited (according to particular bibliographic citation service, e.g., Google Scholar);
- the contexts involved in a citation – i.e., the part P_{citing} of the citing article containing a particular in-text reference pointer and the part P_{cited} of the cited article that is relevant to P_{citing} (useful for some browsing tools of articles and citation services, e.g., *CSIBS* [19] and *CiTalO* [6]).

The *Citation Counting and Context Characterization Ontology*[13] (*C4O*) has been developed to allow the description of the above entities. This ontology enables the characterisation of bibliographic citations in terms of their presence in an article by means of the following classes (shown in Fig. 3):

- class *c4o:InTextReferencePointer*. An in-text reference pointer is a textual device denoting (property *c4o:denotes*) a single bibliographic reference that is embedded in the text of a document within the context of a particular sentence;
- class *c4o:InTextReferencePointerList*. A list containing (through the chain *co:item* and *co:itemContent*) only in-text reference pointers denoting the specific bibliographic references to which the list pertains (property *c4o:pertains*). Such a list cannot contain more than one item containing the same in-text reference pointer;
- class *c4o:SingleReferencePointerList*. Defined as subclass of the previous class, it is an in-text reference pointer list that pertains to exactly one bibliographic reference;
- class *c4o:GlobalCitationCount*. The number of times a work has been cited globally (property *c4o:hasGlobalCountValue*), as determined from a particular bibliographic information source (property *c4o:hasGlobalCountSource*) on a particular date (property *c4o:hasGlobalCountDate*).

C4O provides the ontological structures which allow one to record the number of in-text citations (property *c4o:hasInTextCitationFrequency*, i.e., the number of in-text reference pointers to a single reference in the reference list of the citing article), and also the number of citations a cited entity has received globally (property *c4o:hasGlobalCitationFrequency*), as determined by a bibliographic information resource such as Google Scholar[14], Scopus[15] or Web of Knowledge[16] on a particular date.

Considering again the example in Section 3, we can write a set of assertions according to C4O that describe how many times a reference is used within the citing article and how much the cited article is globally cited (according to Google Scholar):

```
:renear02 a biro:BibliographicReference ;
  c4o:hasInTextCitationFrequency "1"^^xsd:nonNegativeInteger .
```

[13] C4O, the Citation Counting and Context Characterization Ontology: http://purl.org/spar/c4o. The prefix *c4o* refers to entities defined in it.
[14] Google Scholar: http://scholar.google.it.
[15] Scopus: http://www.info.sciverse.com/scopus/.
[16] Web of Knowledge: http://apps.isiknowledge.com.

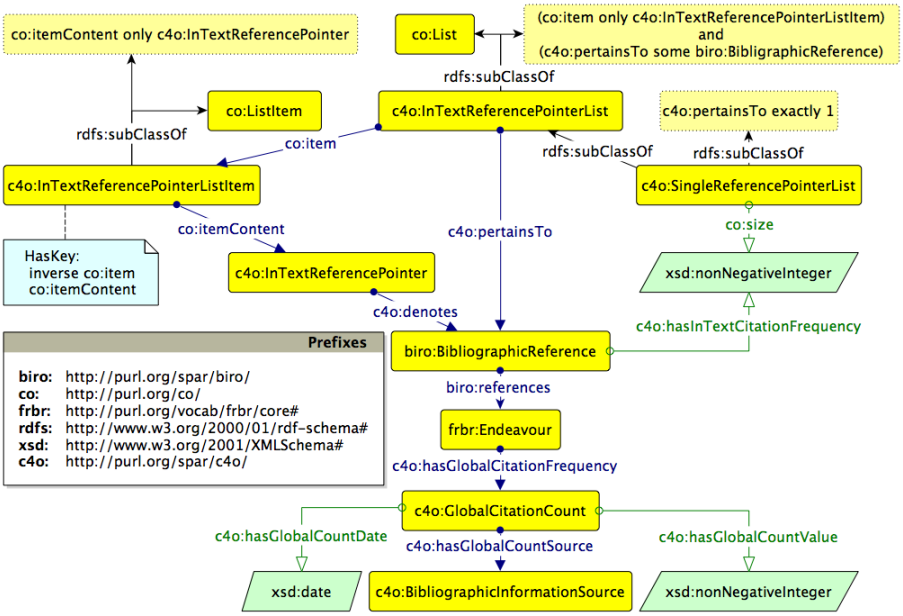

Fig. 3. Graffoo diagram summarising the C4O entities used for counting citations and references.

```
:towards-a-semantics c4o:hasGlobalCitationFrequency [
  a c4o:GlobalCitationCount ;
  c4o:hasGlobalCountDate "2014-03-17"^^xsd:date ;
  c4o:hasGlobalCountSource [ a c4o:BibliographicInformationSource ;
    foaf:homepage <http://scholar.google.com> ] ;
  c4o:hasGlobalCountValue "5"^^xsd:nonNegativeInteger ] .
```

Moreover, C4O enables ontological descriptions of the context where an in-text reference pointer appears in the citing document (modelled as shown in Fig. 4), and allows one to relate that context to relevant textual passages in the cited document.

Considering the previous bibliographic reference example, a possible C4O formalisation of the contexts involved by that citing act is:

```
:intertextual-semantics frbr:part :in-text-renear02 .
:in-text-renear02 a c4o:InTextReferencePointer ;
  c4o:denotes :renear02 ; c4o:hasContext :citation-sentence .
:citation-sentence a doco:Sentence ;
  c4o:hasContent "Renear, Dubin, and Sperberg-McQueen (2002, pp. 121-122)
      proposed a formal semantic approach for structured documents." .
:sentence-in-towards-a-semantics a doco:Sentence ;
  frbr:partOf :towards-a-semantics ;
  c4o:hasContent "Markup semantics are modeled computationally by applying
      knowledge representation technologies to the problem of making those
      structures, relationships, and properties explicit." ;
  c4o:isRelevantTo :citation-sentence .
```

C4O, thus, completes the basic notions behind bibliographic references we introduced in Section 2.

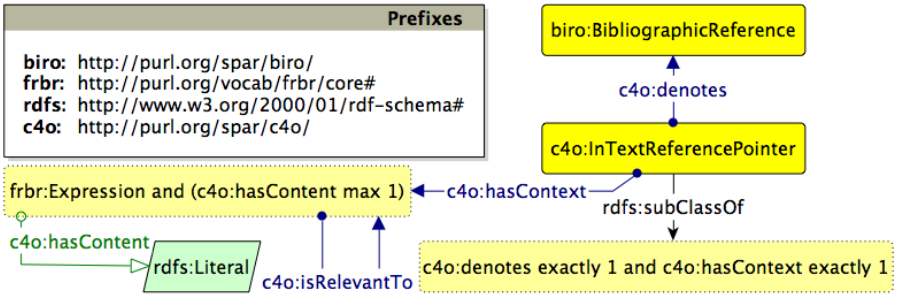

Fig. 4. Graffoo diagram summarising the C4O entities for describing citation contexts.

5 Converting references with REnhancer

In this section we describe a prototype that produces a BiRO-compliant RDF from a textual reference list. We have named the tool *REnhancer*, which stands for ***R****eference list* ***Enhancer***. The tool is freely available online[17] and can be invoked through a Web interface or as a REST API. The input reference list can be provided as formatted in the source article: it can be simply copied and pasted from an article into the text area of the Web interface or passed directly to the Web service. The list could be automatically extracted from PDF or HTML articles by using other tools (e.g., PDFX [4]) and passed to REnhancer.

The tool accepts two optional parameters, i.e., the namespace IRI to use for the generation of identifiers for new generated objects, and the output format for the RDF serialization (RDF/XML, Turtle or N-Triples).

For example, given the following reference list:

1. Di Iorio, A., Nuzzolese, A. G., Peroni, S. (2013). Towards the automatic identification of the nature of citations. In Proceedings of 3rd Workshop on Semantic Publishing (SePublica 2013): 63-74.
2. Garcia Castro, L. J., Berlanga, R., Rebholz-Schuhmann, D., Garcia, A. (2013). Connections across scientific publications based on semantic annotations. In Proceedings of 3rd Workshop on Semantic Publishing (SePublica 2013): 51-62.

the following RDF is returned:

```
:reference-list a biro:ReferenceList ; co:firstItem :reference-list-item-1 ;
  co:item :reference-list-item-1 , :reference-list-item-2 ;
  co:lastItem :reference-list-item-2 ; co:size "2"^^xsd:nonNegativeInteger .
:reference-list-item-1 rdf:type co:ListItem ;
  co:index "1"^^xsd:nonNegativeInteger ; co:itemContent :reference-1 ;
  co:nextItem :reference-list-item-2 .
:reference-1 rdf:type biro:BibliographicReference ;
  dcterms:bibliographicCitation "Di Iorio, A., Nuzzolese, A. G., Peroni, S.
      (2013). Towards the automatic identification of the nature of citations
      . In Proceedings of SePublica 2013: 63-74."^^xsd:string . ...
```

Future releases of the tool will also expand information about authors, publication venue and year.

[17] Renhancer homepage: http://www.cs.unibo.it/~nuzzoles/renhancer.

The following command shows how to use cURL to invoke REnhancer via REST API and how to pass parameters. The textual reference list is passed through the ref-list parameter (by substituting "TEXTUAL REFERENCE LIST" with the actual references) that is mandatory for finalising a request. An optional namespace for new generated items and bibliographic references can be specified through the "namespace" parameter:

```
curl -v -X POST -H "Accept: text/turtle" -d
  namespace="http://foo.org/referenes#"
  --data-urlencode ref-list="TEXTUAL_REFERENCE_LIST"
  http://www.cs.unibo.it/~nuzzoles/renhancer/
```

REnhancer is implemented as a PHP application. The recognition of individual bibliographic references within the reference list is performed by means of regular expressions, which enable to distinguish commonly used syntactic patterns, such as numbered lists, lists based on authors' initials or brackets.

6 Conclusions

The main goal of this paper was to present *BiRO* and *C4O*, two OWL ontologies for the in-depth description of citations in scientific papers. A more accurate and expressive representation of citations makes it possible to better integrate bibliographic information into the Linked Data universe, and to enable sophisticated reasoning and applications. Yet, open datasets about scientific papers and citation networks are already available – such as DBLP and ACM – but they mainly contain bibliographic records and do not describe other details related to citations between papers, such as in-text reference pointers, citation contexts, citation counting, etc. The two ontologies presented here aim at introducing a vocabulary to describe such citation-related entities.

While these ontologies are quite stable, there is a long way to go on the tools for extracting such semantic bibliographic information. In this paper we presented a prototype called REnhancer that takes as input a raw-text list of references and produces a BIRO and C4O translation of these entries. REnhancer is planned to be extended to extract more information and integrated with tools for the automatic extraction of content from PDF and XHTML. A parallel research line we are following and we plan to integrate with REnhancer is on the automatic extraction and characterisation of citations from XML documents.

References

1. Barabucci, G., Di Iorio, A., Peroni, S., Poggi, F., & Vitali, F. (2013). Annotations with EARMARK in practice: a fairy tale. In Proceedings of DH-CASE 2013. DOI: 10.1145/2517978.2517990
2. Ciancarini, P., Di Iorio, A., Nuzzolese, A. G., Peroni, S., & Vitali, F. (2014). Evaluating citation functions in CiTO: cognitive issues. To appear in Proceedings of ESWC 2014. Retrieved April 9, 2014, from http://speroni.web.cs.unibo.it/publications/ciancarini-in-press-evaluating-citation-functions.pdf

3. Ciccarese, P., & Peroni, S. (2013). The Collections Ontology: creating and handling collections in OWL 2 DL frameworks. To appear in Semantic Web – Interoperability, Usability, Applicability. DOI: 10.3233/SW-130121
4. Constantin, A., Pettifer, S., & Voronkov, A. (2013). PDFX: fully-automated PDF-to-XML conversion of scientific literature. In Proceedings of DocEng 2013: 177–180. DOI: 10.1145/2494266.2494271
5. DCMI Usage Board. (2012). DCMI Metadata Terms. DCMI Recommendation, 14 June 2012. Dublin Core Metadata Initiative. Retrieved April 9, 2014, from http://dublincore.org/documents/dcmi-terms/
6. Di Iorio, A., Nuzzolese, A. G., & Peroni, S. (2013). Towards the automatic identification of the nature of citations. In Proceedings of SePublica 2013. http://ceur-ws.org/Vol-994/paper-06.pdf
7. Di Iorio, A., Peroni, S., Poggi, F., Vitali, F., & Shotton, D. (2013). Recognising document components in XML-based academic articles. In Proceedings of DocEng 2013: 181–184. DOI: 10.1145/2494266.2494319
8. D'Arcus, B., & Giasson, F. (2009). Bibliographic Ontology Specification. Specification Document, 4 November 2009. Retrieved April 9, 2014, from http://bibliontology.com/
9. Gangemi, A., Peroni, S., & Vitali, F. (2010). Literal Reification. In Proceedings of WOP 2010: 65–66. http://ceur-ws.org/Vol-671/pat04.pdf
10. Hammond, T. (2008). RDF Site Summary 1.0 Modules: PRISM. Retrieved April 9, 2014, from http://nurture.nature.com/rss/modules/mod_prism.htm
11. Hou, W.-R., Li, M., & Niu, D.-K. (2011). Counting citations in texts rather than reference lists to improve the accuracy of assessing scientific contribution: Citation frequency of individual articles in other papers more fairly measures their scientific contribution than mere presence in reference lists. BioEssays, 33(10): 724–727. DOI: 10.1002/bies.201100067
12. Marcoux, Y., & Rizkallah, E. (2009). Intertextual semantics: A semantics for information design. Journal of the American Society for Information Science and Technology, 60(9): 1895–1906. DOI: 10.1002/asi.21134
13. Peroni, S., & Shotton, D. (2012). FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. Web Semantics: Science, Services and Agents on the World Wide Web, 17: 33–43. DOI: 10.1016/j.websem.2012.08.001
14. Peroni, S., Gangemi, A., & Vitali, F. (2011). Dealing with markup semantics. In Proceedings the 7th International Conference on Semantic Systems: 111–118. DOI: 10.1145/2063518.2063533
15. Peroni, S., Gray, T., Dutton, A., & Shotton, D. (2015). Setting our bibliographic references free: towards open citation data. To appear in Journal of Documentation, 71(2). Retrieved April 9, 2014, from http://speroni.web.cs.unibo.it/publications/peroni-in-press-setting-bibliographic-references.pdf
16. Qazvinian, V., & Radev, D. R. (2010). Identifying Non-explicit Citing Sentences for Citation-based Summarization. In Proceedings of ACL 2010: 555–564.
17. Shotton, D. (2009). Semantic publishing: the coming revolution in scientific journal publishing. Learned Publishing, 22(2): 85–94. DOI: 10.1087/2009202
18. Shotton, D. (2013). Publishing: Open citations. Nature, 502(7471): 295–297. DOI: 10.1038/502295a
19. Wan, S., Paris, C., & Dale, R. (2010). Supporting browsing-specific information needs: Introducing the Citation-Sensitive In-Browser Summariser. Web Semantics: Science, Services and Agents on the World Wide Web, 8(2-3): 196–202. DOI: 10.1016/j.websem.2010.03.002