

Characterising Citations in Scholarly Documents: The CiTalO Framework

Angelo Di Iorio¹, Andrea Giovanni Nuzzolese^{1,2}, and Silvio Peroni^{1,2}

¹ Department of Computer Science and Engineering, University of Bologna, Italy
{diiorio,nuzzoles,essepuntato}@cs.unibo.it

² STLab-ISTC Consiglio Nazionale delle Ricerche, Italy

Abstract. The reasons why an author cites other publications are varied: an author can cite previous works to gain assistance of some sort in the form of background information, ideas, methods, or to review, critique or refute previous works. The problem is that the best possible way to retrieve the nature of citations is very time consuming: one should read article by article to assign a particular characterisation to each citation. In this paper we propose an algorithm, called *CiTalO*, to infer automatically the function of citations by means of Semantic Web technologies and NLP techniques. We also present some preliminary experiments and discuss some strengths and limitations of this approach.

Keywords: CiTO, CiTalO, OWL, WordNet, citation function, semantic publishing.

1 Introduction

The academic community lives on bibliographic citations. First of all, these references are *tools for linking* research. Whenever a researcher writes a paper she/he uses bibliographic references as pointers to related works, to sources of experimental data, to background information, to standards and methods linked to the solution being discussed, and so on. Similarly, citations are *tools for disseminating* research. Not only on academic conferences and journals. Dissemination channels also include publishing platforms on the Web like blogs, wikis, social networks. More recently, semantic publishing platforms are also gaining relevance [15]: they support users in expressing semantic and machine-readable information. From a different perspective, citations are *tools for exploring* research. The network of citations is a source of rich information for scholars and can be used to create new and interesting ways of browsing data. A great amount of research is also being carried on sophisticated visualisers of networks of citations and powerful interfaces allowing users to filter, search and aggregate data. Finally, citations are *tools for evaluating* research – e.g. quantitative metrics on bibliographic references are commonly used for measuring the importance of a journal (the *impact factor*) or the scientific productivity of an author (the *h-index*).

This work begins with the basic assumption that all these activities can be radically improved by exploiting the actual nature of citations. Let us consider

citations as means for evaluating research. Could a paper that is cited many times with negative reviews be given a high score? Could a paper containing several citations of the same research group be given the same score of a paper with heterogeneous citations? How can a paper cited as plagiarism be ranked? These questions can be answered by looking at the nature of the citations, not only their existence. On top of such characterisation, it will also be possible to automatically analyse the pertinence of documents to some research areas, to discover research trends and the structure of communities, to build sophisticated recommenders and qualitative research indicators, and so on.

There are in fact ontologies for describing the nature of citations in scientific research articles and other scholarly works. In the Semantic Web community, the most prominent one is *CiTO* (*Citation Typing Ontology*)¹ [12]. CiTO is written in OWL and is connected to other works in the area of semantic publishing. It is then a very good basis for implementing sophisticated services and for integrating citational data with linked data silos.

The goal of this paper is to present a novel approach to automatically annotate citations with properties defined in CiTO. We present an algorithm and its implementation, called *CiTalO* (from merging the words **CiTO** and *al gorithm*), that takes as input a sentence containing a reference to a bibliographic entity and infers the function of that citation by exploiting Semantic Web technologies and Natural Language Processing (NLP) techniques. The tool is available online at <http://wit.istc.cnr.it:8080/tools/citalo>.

We also present some preliminary tests on a small collection of documents, that confirmed some strengths and weaknesses of such approach. The research direction looks very promising and the CiTalO infrastructure is flexible and extensible. We plan to extend the current set of heuristics and matching rules for a wide practical application of the method.

The paper is then structured as follows. In Section 2 we introduce previous works on classification of citations. In Section 3 we describe our algorithm introducing its structure and presenting the technologies (NLP tools, sentiment analysis procedures, OWL ontologies) we used to develop it. In Section 4 we present the outcome of the algorithm run upon some scientific documents and we discuss those results in Section 5. Finally, in Section 6, we conclude the paper sketching out some future works.

2 Related Works

The automatic analysis of networks of citations is gaining importance in the research community. Copestake *et al.* [4] present an infrastructure called SciBorg that allows one to automatically extract semantic characterisations of scientific texts. In particular, they developed a module for discourse and citation analysis based on the approach proposed by Teufel *et al.* [17] called *Argumentative Zoning* (AZ). AZ provides a procedural mechanism to annotate sentences of an article according to one out of seven classes of a given annotation scheme (i.e.

¹ CiTO: <http://purl.org/spar/cito>

background, own, aim, textual, contrast, basis and other), thus interpreting the intended authors’ motivation behind scientific content and citations.

Teufel *et al.* [18] [19] study the *function* of citations – that they define as “author’s reason for citing a given paper” – and provide a categorisation of possible citation functions organised in twelve classes, in turn clustered in *Negative*, *Neutral* and *Positive* rhetorical functions. In addition, they describe the outcomes of some tests involving hundreds of article in computational linguistics (stored as XML files), several human annotators and a machine learning (ML) approach for the automatic annotation of citation functions. Their approach is quite promising; however the agreement between humans (i.e. $K = 0.72$) is still higher than the one between the humans and the ML approach (i.e. $K = 0.57$).

Jorg [9] introduces an analysis of the ACL Anthology Networks² and identifies one hundred fifty *cue verbs*, i.e. verbs usually used to carry important information about the nature of citations: *based on, outperform, focus on, extend*, etc. She maps cue verbs to classes of citation functions according to the classification provided by Moravcsik *et al.* [10] and makes the bases to the development of a formal citation ontology. This works actually represent one of the sources of inspiration of *CiTO* (the *Citation Typing Ontology*) developed by Peroni *et al.* [12], which is an ontology that permits the motivations of an author when referring to another document to be captured and described by using Semantic Web technologies such as RDF and OWL.

Closely related to the annotation of citation functions, Athar [1] proposes a sentiment-analysis approach to citations, so as to identify whether a particular act of citing was done with positive (e.g. praising a previous work on a certain topic) or negative intentions (e.g. criticising the results obtained through a particular method). Starting from empirical results Athar *et al.* [2] expand the above study and show how the correct sentiment (in particular, a negative sentiment) of a particular citation usually does not emerge from the citation sentence – i.e. the sentence that contains the actual pointer to the bibliographic reference of the cited paper. Rather, it actually becomes evident in the last part of the *context window*³ [14] in consideration.

Hou *et al.* [8] use an alternative approach to understand the importance (seen as a form of positive connotation/sentiment) of citations: the citation counting in text. Paraphrasing the authors, the idea is that the more a paper is cited within a text, the more its scientific contribution is significative.

3 Our Approach

In this section, we introduce CiTalO, a tool that infers the function of citations by combining techniques of ontology learning from natural language, sentiment-analysis, word-sense disambiguation, and ontology mapping. These techniques

² ACL Anthology Network: <http://clair.eecs.umich.edu/aan/index.php>

³ The *context window* [14] of a citation is a chain of sentences implicitly referring to the citation itself, which usually starts from the citation sentence and involves few more subsequent sentences where that citation is still implicit [3].

are applied in a pipeline whose input is the textual context containing the citation and the output is a one or more properties of CiTO [12].

The overall CiTaLO schema is shown in Fig. 1. It was inspired by Gangemi *et al.*'s work [7], in which a similar pipeline was used with good results for automatically typing DBpedia resources by analysing corresponding Wikipedia abstracts. Five steps (described below) compose the architecture, and each one is implemented as a pluggable OSGi component [11] over a Pipeline Manager that coordinates the process.

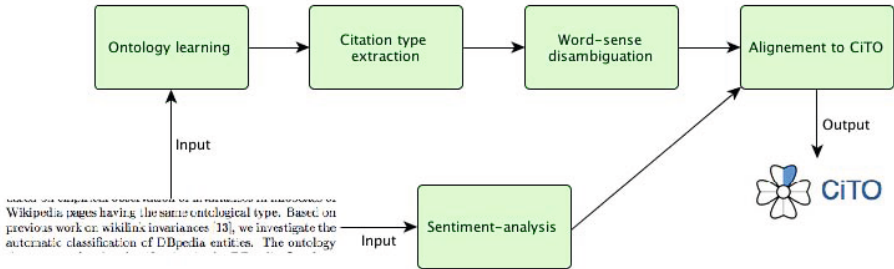


Fig. 1. Pipeline used by CiTaLO. The input is the textual context in which the citation appears and the output is a set of properties of the CiTO ontology.

In order to detail the components of CiTaLO we will discuss how the algorithm works on the following sample sentence: “It extends the research outlined in earlier work X.”, where “X” is the cited work.

Sentiment-analysis to Gather the Polarity of the Citational Function.

The aim of the sentiment-analysis in our context is to capture the sentiment polarity emerging from the text in which the citation is included. The importance of this step derives from the classification of CiTO properties according to three different polarities, i.e., positive, neuter and negative. This means that being able to recognise the polarity behind the citation would restrict the set of possible target properties of CiTO to match. We are currently using AlchemyAPI⁴, a suite of sentiment-analysis and NLP tools that exposes its services through HTTP REST interfaces. The output returned by this component with respect to our example is a positive polarity.

Ontology Extraction from the Textual Context of the Citation. The first mandatory step of CiTaLO consists of deriving a logical representation of the sentence containing the citation. The ontology extraction is performed by using FRED [13], a tool for ontology learning based on discourse representation theory, frames and ontology design patterns. Such an approach follows the one proposed by Gangemi *et al.* [7], which exploited FRED for automatically typing DBpedia entities. The transformation of the sentence into a logical form allows us

⁴ AlchemyAPI: <http://www.alchemyapi.com>

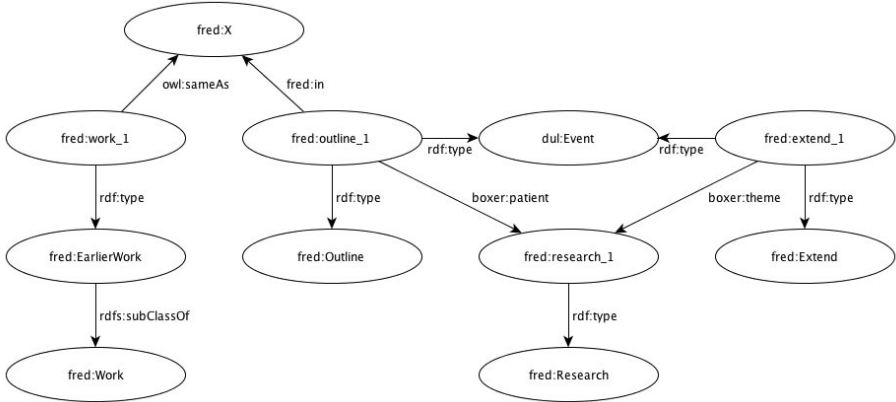


Fig. 2. FRED result for “It extends the research outlined in earlier work X”

to recognise graph-based heuristics in order to detect possible types of functions of the citation. The output of FRED on our example is shown in Fig. 2. FRED recognises two events, i.e., *Outline* and *Extend*, and the cited work *X* is typed as *EarlierWork* that is subclass of *Work*.

Citation Type Extraction through Pattern Matching. The second step consists of extracting candidate types for the citation, by looking for patterns in the FRED result. In order to collect these types we have designed ten graph-based heuristics and we have implemented them as SPARQL queries. The pattern matcher tries to apply all the patterns, which are namely:

```

SELECT ?type WHERE {?subj ?prop fred:X ; a ?type}
SELECT ?type WHERE {?subj ?prop fred:X ; a ?typeTmp .
  ?typeTmp rdfs:subClassOf+ ?type}
SELECT ?type WHERE {?subj a dul:Event , ?type .
  FILTER(?type != dul:Event)}
SELECT ?type WHERE {?subj a dul:Event , ?typeTmp .
  ?typeTmp rdfs:subClassOf+ ?type.FILTER(?type != dul:Event)}
SELECT ?type WHERE {?subj a dul:Event ;
  boxer:theme ?theme . ?theme a ?type}
SELECT ?type WHERE {?subj a dul:Event ; boxer:theme ?theme .
  ?theme a ?typeTmp . ?typeTmp rdfs:subClassOf+ ?type}
SELECT ?type WHERE {?subj a dul:Event ;
  boxer:patient ?patient . ?patient a ?type}
SELECT ?type WHERE {?subj a dul:Event ; boxer:patient ?pat .
  ?pat a ?typeTmp . ?typeTmp rdfs:subClassOf+ ?type}
SELECT ?type WHERE {?subj a dul:Event ; boxer:patient ?pat .
  ?pat ?prop ?any . ?any a ?type}
SELECT ?type WHERE {?subj a dul:Event ; boxer:patient ?pat .
  ?pat ?prop ?any . ?any a ?typeTmp .
  ?typeTmp rdfs:subClassOf+ ?type}

```

Applying these patterns to the example the following candidate types are found: *Outline*, *Extend*, *EarlierWork*, *Work*, and *Research*. The current set of patterns is quite simple and incomplete. We are investigating new patterns and we are continuously updating the catalogue.

Word-sense Disambiguation. In order to gather the sense of candidate types we need a word-sense disambiguator. For this purpose we used IMS [20], a tool based on linear support vector machines. The disambiguation is performed with respect to OntoWordNet [6] (the OWL version of WordNet) and produces a list of synsets for each candidate type. The following disambiguations are returned on our example: (i) *Extend* is disambiguated as `own:synset-prolong-verb-1`, (ii) *Outline* as `own:synset-delineate-verb-3`, (iii) *Research* as `own:synset-research-noun-1`, (iv) *EarlierWork* and *Work* as `own:synset-work-noun-1`.

The output of this step can also be extended by adding *proximal synsets*, i.e. synsets that are not directly returned by IMS but whose meaning is close to those found while disambiguating. To do so, we use the RDF graph of proximity introduced in [7].

Alignment to CiTO. The final step consists of assigning CiTO types to citations. We use two ontologies for this purpose: *CiTOFunctions* and *CiTO2Wordnet*. The *CiTOFunctions* ontology⁵ classifies each CiTO property according to its factual and positive/neutral/negative rhetorical functions, using the classification proposed by Peroni *et al.* [12].

*CiTO2Wordnet*⁶ maps all the CiTO properties defining citations with the appropriate Wordnet synsets (as expressed in OntoWordNet). This ontology was built in three steps:

- *identification step.* We identified all the Wordnet synsets related to each of the thirty-eight sub-properties of *cites* according to the verbs and nouns used in property labels (i.e. *rdfs:label*) and comments (i.e. *rdfs:comment*) – for instance, the synsets *credit#1*, *accredit#3*, *credit#3*, *credit#4* refers to the property *credits*;
- *filtering step.* For each CiTO property, we filtered out all those synsets of which the *gloss*⁷ is not aligned with the natural language description of the property in consideration – for instance, the synset *credit#3* was filtered out since the gloss “accounting: enter as credit” means something radically different to the CiTO property description “the citing entity acknowledges contributions made by the cited entity”;
- *formalisation step.* finally, we linked each CiTO property to the related synsets through the property *skos:closeMatch*. An example in Turtle is: `cito:credits skos:closeMatch own:synset-credit-verb-1`.

⁵ *CiTOFunctions*:<http://www.essepuntato.it/2013/03/cito-functions>

⁶ *CiTO2Wordnet* ontology:<http://www.essepuntato.it/2013/03/cito2wordnet>

⁷ In Wordnet, the *gloss* of a synset is its natural language description.

The final alignment to CiTO is performed through a SPARQL CONSTRUCT query that uses the output of the previous steps, the polarity gathered from the sentiment-analysis phase, OntoWordNet and the two ontologies just described. In the case of empty alignments, the CiTO property *citesForInformation* is returned as base case. In the example, the property *extends* is assigned to the citation.

4 Testing and Evaluation

The test consisted of comparing the results of CiTalO with a human classification of the citations. The test bed we used for our experiments includes some scientific papers (written in English) encoded in XML DocBook, containing citations of different types. The papers were chosen among those published in the proceedings of the Balisage Conference Series. In particular, we automatically extracted citation sentences, through an XSLT document⁸, from all the papers published in the seventh volume of Balisage Proceedings, which are freely available online⁹. For our test, we took into account only those papers for which the XSLT transform retrieved at least one citation (i.e. 18 papers written by different authors). The total number of citations retrieved was 377, for a mean of 20.94 citations per paper. Notice that the XSLT transform was quite simple at that stage. It basically extracted the *citation sentence* around a citation (i.e. the sentence in which that citation is explicitly used), preparing data for the actual CiTalO pipeline.

We first filtered all the citation sentences from the selected articles, and then we annotated them manually using the CiTO properties. Since the annotation of citation functions is actually an hard problem to address – it requires an interpretation of author intentions – we mark only the citations that are accompanied by verbs (*extends*, *discusses*, etc.) and/or other grammatical structures (*uses method in*, *uses data from*, etc.) carrying explicitly a particular citation function. We considered that rule as a strict guideline as also suggested by Teufel *et al.* [18].

We marked 106 citations of out the 377 originally retrieved, obtaining at least one representative citation for each of the 18 paper used (with a mean of 5.89 citations per paper). We used 21 CiTO properties out of 38 to annotate all these citations, as shown in Table 1.

Interesting similarities can be found between such a classification and the results of Teufel *et al.* [19]. In this paper, the neutral category *Neut* was used for the majority of annotations by humans; similarly the most neutral CiTO property, *citesForInformation*, was the most prevalent function in our dataset too. The second most used property was *usedMethodIn* in both analyses.

⁸ Available at <http://www.essepuntato.it/2013/sepublica/xslt>

⁹ Proceedings of Balisage 2011:<http://balisage.net/Proceedings/vol7/cover.html>

Table 1. The way we marked the citations within the 18 Balisage papers.

# Citations	CiTO property
53	citesForInformation
15	usesMethodIn
12	usesConclusionsFrom
11	obtainsBackgroundFrom
8	discusses
4	citesAsRelated, extends, includesQuotationFrom, citesAsDataSource, obtainsSupportFrom
< 4	credits, critiques, useConclusionsFrom, citesAsAuthority, usesDataFrom, supports, updates, includesExcerptFrom, includeQuotationForm, citesAsRecommendedReading, corrects

Table 2. The number of true positives, false positives and false negatives returned by running CiTalO with the eight different configurations

Configuration	TP	FP	FN
Filtered (with or without Sentiment)	47	88	59
Filtered + Proximity	40	137	66
Filtered + Proximity + Sentiment	41	136	65
All (with or without Sentiment)	52	114	54
All + Proximity (with or without Sentiment)	45	174	64

We run CiTalO on these data (i.e. 106 citations in total) and compared results with our previous analysis¹⁰. We also tested eight different configurations of CiTalO, corresponding to all possible combinations of three options:

- activating or deactivating the sentiment-analysis module;
- applying or not the proximal synsets¹¹ to the word-disambiguation output;
- using the CiTO2Wordnet ontology as described in Section 3, or an extended version that also includes all the discarded synsets during the filtering step.

The number of *true positives* (TP), *false positives* (FP) and *false negatives* (FN) obtained comparing CiTalO outcomes with our annotations are shown in Table 2.

¹⁰ All the source materials we used for the test is available online at <http://www.essepuntato.it/2013/sepublica/test>. Note that a comparative evaluation with other approaches, such as Teufel’s, was not feasible at this stage since input data and output categories were heterogeneous and were not directly comparable.

¹¹ We used the same the RDF graph of proximal synsets introduced in [7].

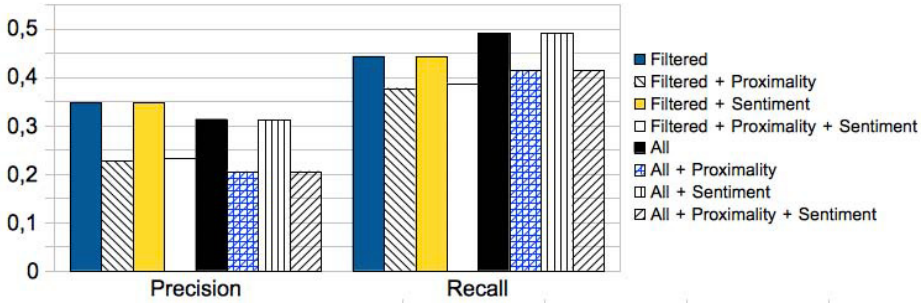


Fig. 3. Precision and recall according to the different configurations used

We calculated the precision – i.e. $TP / (TP + FP)$ – and the recall – i.e. $TP / (TP + FN)$ – obtained by using each configuration. As shown in Fig. 3, *Filtered* and *Filtered+Sentiment* had the best precision (i.e. 0.348) and the second recall (i.e. 0.443), while *All* and *All+Sentiment* had the second precision (i.e. 0.313) and the best recall (i.e. 0.491).

There is no configuration that emerges as the best one from these data. They rather suggest an hybrid approach that also takes into account some of the discarded synsets. It is evident that the worst configurations were those that took into account all the proximal synsets. It looks that the more synsets CiTaLO uses, the less the citation functions retrieved conform to humans’ annotations.

In general, the values of precision and recall of our experiments are quite low. However, our preliminary tests aimed at defining a baseline for future developments of our approach, more than a definitive evaluation of CiTaLO effectiveness.

5 Limitations and Future Research Directions

In this section we discuss some limitations and possible improvements of CiTaLO outlined by the tests and that we plan to address in future releases of the tool.

Coverage of CiTO Properties. The manual annotation process highlighted that CiTO properties do not cover all the citation scenarios addressed in the experiment. For instance, let us consider the following sentence from [16]: “*We speculate that some Goddag-based structure analogous to the multi-coloured trees of [Jagadish et al. 2004] may be a useful solution*”.

The verb *speculate* used above is very specific and refers to synsets that are not included in the mapping defined in the CiTO2Wordnet ontology. This kind of citation is not explicitly mentioned in CiTO neither. The same happens for citations – usually they are introduced by modal verbs – that suggest a work as *potential solution* for an issue related to the paper in consideration, for instance the aforementioned sentence and the following one (again from [16]): “*Mechanisms like Trojan Horse markup ([DeRose 2004], [Bauman 2005]) can be used to serialize discontinuous elements*”.

What is needed here is an accurate analysis of citations in papers so as to suggest some extensions to CiTO itself. Towards this direction, a good

starting point is to use Jorg’s previous work on cue verbs [9], where she listed one hundred-fifty verbs that are typically used in citations within scientific articles.

Noise of Proximity Synsets. The diagram in Fig. 3 clearly shows that using proximity synsets decreased both precision and recall. One would expect, on the other hand, that a larger set of synsets produced better results.

This depends on the number of *citesAsInformation* retrieved by CiTalO (re-mind that *citesAsInformation* is assigned when no further CiTO property is identified). Let us consider the case of *Filtered: citesForInformation* was assigned correctly 42 times out of 47 occurrences¹², while using *Filtered+Proximity* the same property was detected only 31 times and other more specific CiTO properties were assigned instead. The problem is that those assignments are not correct, as they derive from proximal synsets that are actually too far from the ones being processed in CiTO2Wordnet. These synsets should not be considered or, at least, should be given less importance than others that are closer to the ones in CiTO2Wordnet. For future releases of CiTalO, in fact, we plan to use proximal synsets distance in order to reduce such a noise.

Matching Synsets and Compound-word Properties. The current CiTalO alignment between synsets and CiTO properties does not work properly with properties described by compound words, such as *useMethodIn*. In fact, CiTalO returns a match whether one of the synsets of the compound words matches with a CiTO property. For instance, let us consider the following sentence (from [16]): “*Later versions of the TEI Guidelines [ACH/ACL/ALLC 1994] define more powerful methods of encoding discontinuity*”.

CiTalO returns the property *usesMethodIn* since one of the related synsets of that property, i.e. *synset:method-noun-1*, was actually found. This output is not correct, since that property should be returned only if there exists evidence that the current work *uses* (a term that is actually missed from that sentence) a particular *method* from another article, while here it seems not to be the case. Future version of CiTalO must take into account these scenarios too.

Identification of the Context Window of Citations. In our experiments, we always used the citation sentence as input of CiTalO. However, as previously noticed by Athar *et al.* [2], the actual intended sentiment and motivation of a citation is not always present in the citation sentence. It may be explicit in some other sentences close to the citation sentence and can refer implicitly to the cited work (through authors’ names, project’s name, pronouns, etc.). The identification of the right citational *context window* [14] is a complex issue that should be addressed to improve the effectiveness of CiTalO.

Identification of Implicit Citations. The identification of *implicit citations* [3] is another issue related to the one being discussed. Let us consider some sentences of a paragraph from [16]: “*XCONCUR and similar mechanisms [Hilbert/*

¹² The other citation functions retrieved are: *citesAsRecommendedReading*, *uses-DataFrom*, *citesAsDataSource*, *extends* and *usesMethodIn* – all of them used just one time within the true positive set.

Schonefeld/Witt 2005] already incorporate the containment/dominance distinction to a certain degree. [...] And like non-concurrent XML, XCONCUR has no conception of discontinuous elements”.

While in the first sentence, it seems that the authors want to praise with a positive connotation the work done by others (i.e. XCONCUR), in the latter sentence they criticise them. The “XCONCUR” in the latter sentence actually represents an implicit citation of the reference contained in the former sentence and, in this case, delimits also the context window of the citation itself. Detecting such scenarios is a further refinement that can improve CiTalO results.

Using Rhetoric Structures. According to Teufel *et al.* [18], recognising implicit citations and context windows “is often not informative enough for the searcher to infer the relation” of citations. Further information can be given by also identifying the rhetorical function of the entire paragraph or section in which the citation appears. For instance, all the references in the “related works” section are usually used to indicate related articles (i.e. *citesAsRelated*) to the topic under consideration, while citations in the introduction present background information (i.e. *obtainsBackgroundFrom*) of the field in which the work described in the article is placed. We are thinking to apply existing techniques of automatic recognition of document structures, e.g. that proposed by Di Iorio *et al.* [5], to retrieve the rhetoric function of sections in scientific articles and integrate such analysis with CiTalO.

6 Conclusions

The implementation of CiTalO is still at an early stage; current experiments are admittedly not enough to fully validate this approach. However, the overall approach is very open to incremental refinements. The goal of this work, in fact, was to build such a modular architecture, to perform some exploratory experiments and to identify issues and possible developments of our approach. We are currently working to include a mechanism for the automatic identification of *context windows* of citations given an input article and to improve *patterns’ matching* phases in CiTalO. In addition, we plan to perform exhaustive tests with a larger set of documents and users.

References

1. Athar, A.: Sentiment Analysis of Citations using Sentence Structure-Based Features. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 81–87 (2011)
2. Athar, A., Teufel, S.: Context-Enhanced Citation Sentiment Detection. In: Proceedings of Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, pp. 597–601 (2012)
3. Athar, A., Teufel, S.: Detection of implicit citations for sentiment detection. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pp. 18–26 (2012)

4. Copestake, A., Corbett, P., Murray-Rust, P., Rupp, C.J., Siddharthan, A., Teufel, S., Waldron, B.: An architecture for language processing for scientific text. In: Proceedings of the UK e-Science All Hands Meeting (2006)
5. Di Iorio, A., Peroni, S., Poggi, F., Vitali, F.: A first approach to the automatic recognition of structural patterns in XML documents. In: Proceedings of the 2012 ACM Symposium on Document Engineering, pp. 85–94 (2012), doi:10.1145/2361354.2361374
6. Gangemi, A., Navigli, R., Velardi, P.: The OntoWordNet Project: Extension and Atomization of Conceptual Relations in WordNet. In: Meersman, R., Schmidt, D.C. (eds.) CoopIS/DOA/ODBASE 2003. LNCS, vol. 2888, pp. 820–838. Springer, Heidelberg (2003)
7. Gangemi, A., Nuzzolese, A.G., Presutti, V., Draicchio, F., Musetti, A., Ciancarini, P.: Automatic Typing of DBpedia Entities. In: Proceedings of the 11th International Semantic Web Conference, pp. 65–81 (2012), doi:10.1007/978-3-642-35176-1_5
8. Hou, W., Li, M., Niu, D.: Counting citations in texts rather than reference lists to improve the accuracy of assessing scientific contribution. *BioEssays* 33(10), 724–727 (2011), doi:10.1002/bies.201100067
9. Jorg, B.: Towards the Nature of Citations. In: Poster Proceedings of the 5th International Conference on Formal Ontology in Information Systems (2008)
10. Moravcsik, M.J., Murugesan, P.: Some Results on the Function and Quality of Citations. *Social Studies of Science* 5(1), 86–92 (1975)
11. OSGi Alliance, OSGi service platform, release 3. IOS Press, Inc. (2003)
12. Peroni, S., Shotton, D.: FaBiO and CiTO: ontologies for describing bibliographic resources and citations. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web* 17, 33–43 (2012), doi:10.1016/j.websem.2012.08.001
13. Presutti, V., Draicchio, F., Gangemi, A.: Knowledge extraction based on discourse representation theory and linguistic frames. In: ten Teije, A., Völker, J., Handschuh, S., Stuckenschmidt, H., d’Acquin, M., Nikolov, A., Aussenac-Gilles, N., Hernandez, N. (eds.) EKAW 2012. LNCS, vol. 7603, pp. 114–129. Springer, Heidelberg (2012)
14. Qazvinian, V., Radev, D.R.: Identifying non-explicit citing sentences for citation-based summarization. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 555–564 (2010)
15. Shotton, D.: Semantic publishing: the coming revolution in scientific journal publishing. *Learned Publishing* 22(2), 85–94 (2009), doi:10.1087/2009202
16. Sperberg-McQueen, C.M., Huitfeldt, C.: Markup Discontinued: Discontinuity in TexMeCs, Goddag structures, and rabbit/duck grammars. In: Proceedings of Balisage: The Markup Conference 2008 (2008), doi:10.4242/BalisageVoll1.Sperberg-McQueen01
17. Teufel, S., Carletta, J., Moens, M.: An annotation scheme for discourse-level argumentation in research articles. In: Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics, pp. 110–117 (1999)
18. Teufel, S., Siddharthan, A., Tidhar, D.: Automatic classification of citation function. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pp. 103–110 (2006)
19. Teufel, S., Siddharthan, A., Tidhar, D.: An annotation scheme for citation function. In: Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue, pp. 80–87 (2009)
20. Zhong, Z., Ng, H.T.: It Makes Sense: A wide-coverage word sense disambiguation system for free text. In: Proceedings of the ACL 2010 System Demonstrations, pp. 78–83 (2010)