Semantic representation of scientific literature: bringing claims, contributions and named entities onto the Linked Open I

Published on semanticsoftware.info (https://www.semanticsoftware.info)

Semantic representation of scientific literature: bringing claims, contributions and named entities onto the Linked Open Data cloud

Submitted by [bahar](#) [1] on Thu, 2016-01-07 16:09

- [Linked Open Data](#) [2]
- [Literature Management](#) [3]
- [natural language processing](#) [4]
- [Semantic Publishing](#) [5]
- [Semantic Publishing](#) [6]
- [Semantic Publishing](#) [7]
- [Semantic Web](#) [8]
- [Semantic Web](#) [9]
- [Semantic Computing](#) [10]
- [NLP](#) [11]
- [Text Mining](#) [12]

Title	Semantic representation of scientific literature: bringing claims, contributions and named entities onto the Linked Open Data cloud
Publication Type	Journal Article
Year of Publication	2015
Authors	[Sateli, B.](#) [13], and [R. Witte](#) [14]
Refereed Designation	Refereed
Editors	[Sumner, T.](#) [15]
Journal	PeerJ Computer Science
Volume	1
Issue	e37
Date Published	12/2015
Keywords	[natural language processing](#) [16], [Semantic Publishing](#) [17], [Semantic Web](#) [18]

Abstract	Motivation. Finding relevant scientific literature is one of the essential tasks researchers are facing on a daily basis. Digital libraries and web information retrieval techniques provide rapid access to a vast amount of scientific literature. However, no further automated support is available that would enable fine-grained access to the knowledge 'stored' in these documents. The emerging domain of Semantic Publishing aims at making scientific knowledge accessible to both humans and machines, by adding semantic annotations to content, such as a publication's contributions, methods, or application domains. However, despite the promises of better knowledge access, the manual annotation of existing research literature is prohibitively expensive for wide-spread adoption. We argue that a novel combination of three distinct methods can significantly advance this vision in a fully-automated way: (i) Natural Language Processing (NLP) for Rhetorical Entity (RE) detection; (ii) Named Entity (NE) recognition based on the Linked Open Data (LOD) cloud; and (iii) automatic knowledge base construction for both NEs and REs using semantic web ontologies that interconnect entities in documents with the machine-readable LOD cloud.

Results. We present a complete workflow to transform scientific literature into a semantic knowledge base, based on the W3C standards RDF and RDFS. A text mining pipeline, implemented based on the GATE framework, automatically extracts rhetorical entities of type Claims and Contributions from full-text scientific literature. These REs are further enriched with named entities, represented as URIs to the linked open data cloud, by integrating the DBpedia Spotlight tool into our workflow. Text mining results are stored in a knowledge base through a flexible export process that provides for a dynamic mapping of semantic annotations to LOD vocabularies through rules stored in the knowledge base. We created a gold standard corpus from computer science conference proceedings and journal articles, where Claim and Contribution sentences are manually annotated with their respective types using LOD URIs. The performance of the RE detection phase is evaluated against this corpus, where it achieves an average F-measure of 0.73. We further demonstrate a number of semantic queries that show how the generated knowledge base can provide support for numerous use cases in managing scientific literature.

Availability. All software presented in this paper is available under open source licenses at http://www.semanticsoftware.info/semantic-scientific-literature-peerj-20... [19]. Development releases of individual components are additionally available on our GitHub page at https://github.com/SemanticSoftwareLab [20].

Attachment	Size
peerj-cs-37.pdf [23]	8.69 MB

Source URL (retrieved on *2026-01-08 06:36*):
https://www.semanticsoftware.info/biblio/semantic-representation-scientific-literature-peerj-compsci-2015

Semantic representation of scientific literature: bringing claims, contributions and named entities onto the Linked Open I

Published on semanticsoftware.info (https://www.semanticsoftware.info)

Links:

[1] https://www.semanticsoftware.info/users/bahar

[2] https://www.semanticsoftware.info/taxonomy/term/418

[3] https://www.semanticsoftware.info/taxonomy/term/391

[4] https://www.semanticsoftware.info/category/blog-tags/natural-language-processing

[5] https://www.semanticsoftware.info/taxonomy/term/419

[6] https://www.semanticsoftware.info/taxonomy/term/390

[7] https://www.semanticsoftware.info/category/blog-tags/semantic-publishing

[8] https://www.semanticsoftware.info/category/blog-tags/semantic-web

[9] https://www.semanticsoftware.info/category/topic/semantic-web

[10] https://www.semanticsoftware.info/category/topic/semantic-computing

[11] https://www.semanticsoftware.info/category/topic/nlp

[12] https://www.semanticsoftware.info/category/topic/text-mining

[13] https://www.semanticsoftware.info/biblio/author/73

[14] https://www.semanticsoftware.info/biblio/author/1

[15] https://www.semanticsoftware.info/biblio/author/161

[16] https://www.semanticsoftware.info/biblio/keyword/16

[17] https://www.semanticsoftware.info/biblio/keyword/104

[18] https://www.semanticsoftware.info/biblio/keyword/2

[19] http://www.semanticsoftware.info/semantic-scientific-literature-peerj-2015-supplements

[20] https://github.com/SemanticSoftwareLab

[21] https://peerj.com/articles/cs-37/

[22] http://dx.doi.org/10.7717/peerj-cs.37

[23] https://www.semanticsoftware.info/system/files/peerj-cs-37.pdf