

# Rhetector: Automatic Detection of Rhetorical Entities in Scientific Literature

- [Semantic Publishing](#)
- [Semantic Publishing](#)
- [Text Mining](#)

toc\_collapse=0; Table of Contents

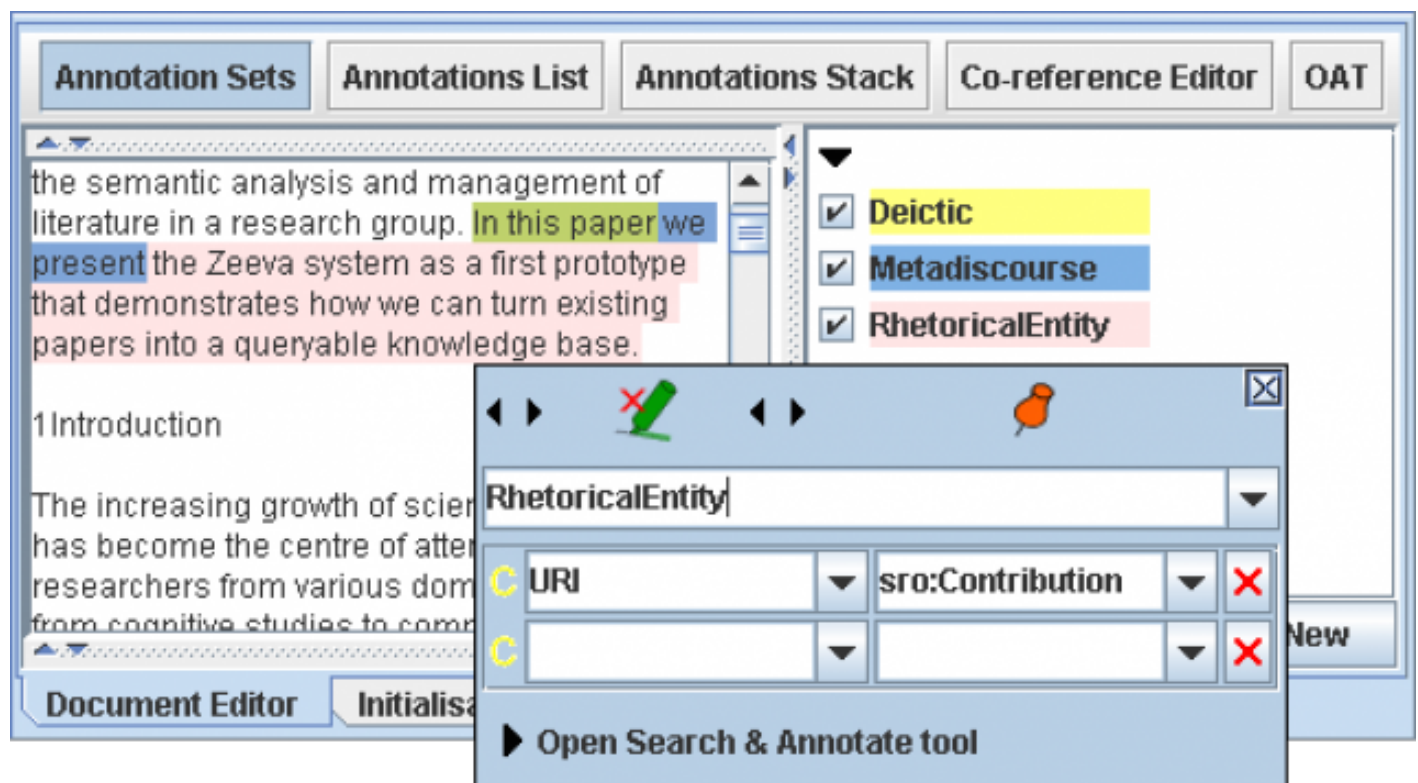
- [1. What are Rhetorical Entities?](#)
- [2. Features](#)
- [3. Download](#)
- [4. More Information and citation](#)
- [5. License](#)
- [6. Version history](#)
- [7. Feedback](#)

Our *Rhetector* component is a [GATE](#) plugin for the automatic detection of Rhetorical Entities (REs) in scientific literature. For background information on the design and application of REs, please read our paper [1]: [Sumner, T.](#) (Eds.), [Sateli, B.](#), and [R. Witte](#), "Semantic representation of scientific literature: bringing claims, contributions and named entities onto the Linked Open Data cloud", *PeerJ Computer Science*, vol. 1, no. e37 PeerJ, 12/2015.

## 1. What are Rhetorical Entities?

In the context of scientific literature, Rhetorical Entities (REs) are spans of text (sentences, passages, sections, etc.) in a document, where authors convey their findings, like Claims or Arguments, to the readers. REs are usually situated in certain parts of a document, depending on their role. For example, the authors' Claims are mentioned in the Abstract, Introduction or Conclusion section of a paper, and seldom in the Background. This conforms with the researchers' habit in both reading and writing scientific articles. Verbatim extraction of REs from text helps to efficiently allocate the attention of humans when reading a paper, as well as improving retrieval mechanisms by finding documents based on their REs (e.g., "Give me all papers with implementation details") [1], [2].

## 2. Features



### Example annotations generated by Rhetector on a document

For each detected RE, an annotation “RhetoricalEntity” is added to the document. Based on the grammatical structure of the RE, it is classified and mapped onto existing concepts on the Linked Open Data (LOD) cloud. The fully-qualified URI of the RE type is stored as the value of the “URI” feature of each annotation.

## 3. Download

Rhetector is available for direct installation through our GATE update site. If you are not familiar with GATE's Plugin Manager, please follow the installation steps in the [documentation](#) (this is the same file as included in the distribution in the doc/ folder) to install the plugin. The download package includes the component's source code, the user guide documentation and a demo pipeline. You can download the [install package](#) manually (but the recommended way of installation is to use the GATE plugin manager through the GATE Developer GUI).

You can check out the latest development version of Rhetector from our public [GitHub repository](#). A continuous integration build of this repository is available on our [Jenkins server](#).

## 4. More Information and citation

For more information on the automatic extraction of REs, please refer to our latest publication [\[1\]](#). If you use our component, we would appreciate a citation of our paper.

## 5. License

The Rhetector component and resources are published under the [GNU Lesser General Public License v3 \(LGPL3\)](#).

## 6. Version history

First release was v1.0 (10.08.2015)

## 7. Feedback

For questions, comments, etc., please use the [Forum](#).

---

## References

1. Sumner, T. (Eds.), Sateli, B., and R. Witte. "Semantic representation of scientific literature: bringing claims, contributions and named entities onto the Linked Open Data cloud". *PeerJ Computer Science*, vol. 1, no. e37 PeerJ, 12/2015.
2. Sateli, B., and R. Witte. "What's in this paper? Combining Rhetorical Entities with Linked Open Data for Semantic Literature Querying". *Semantics, Analytics, Visualisation: Enhancing Scholarly Data (SAVE-SD 2015)*, Florence, Italy : ACM, pp. 1023–1028, 05/2015.



Except where otherwise noted, all original content on this site is copyright by its author and licensed under a [Creative Commons Attribution-Share Alike 2.5 Canada License](#).

Source URL (retrieved on 2025-12-11 10:07): <https://www.semanticssoftware.info/rhetector>