# Classifying Research Papers with the Computer Science Ontology

A. Salatino, T. Thanapalasingam, A. Mannocci, F. Osborne, E. Motta

## The Computer Science Ontology

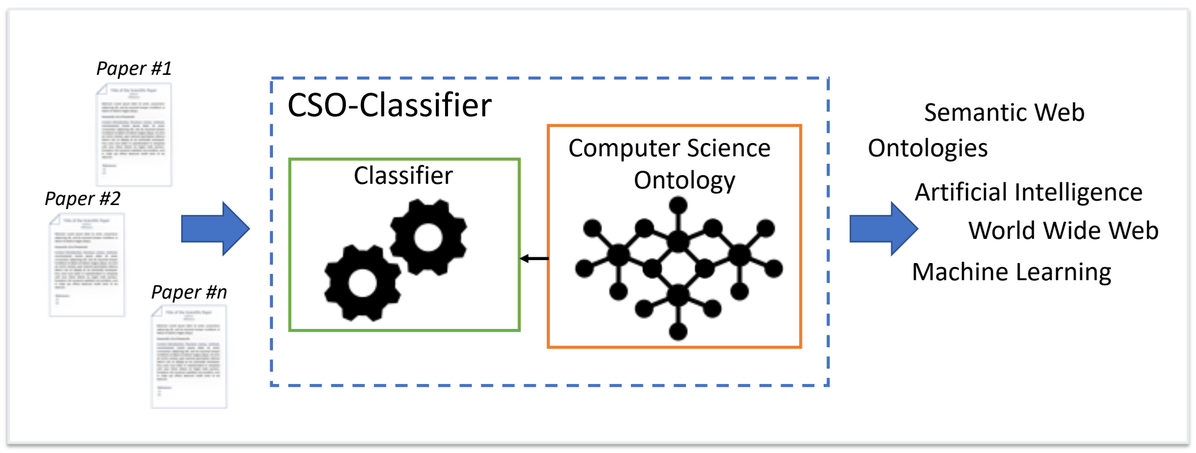
The **Computer Science Ontology** (CSO) is an ontology of research areas that was automatically generated using the Klink-2 algorithm [1] on a dataset of 16 million publications, mainly in the field of Computer Science. Ontologies like CSO are invaluable tools for: i) making sense of the research dynamics, ii) classifying publications, iii) identifying research communities, and iv) forecasting research trends [2].

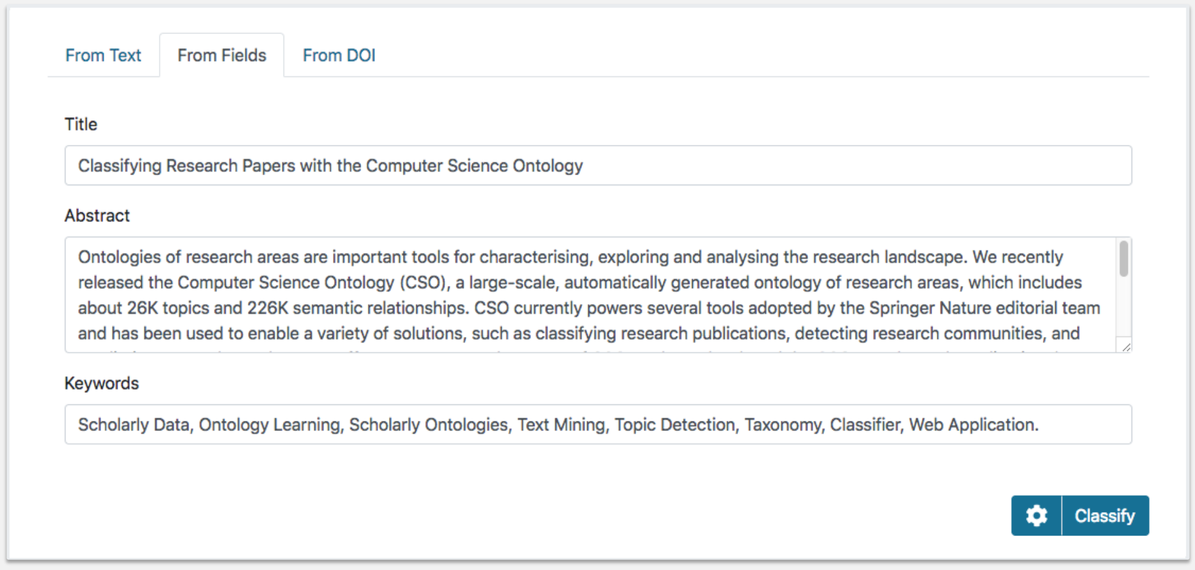
CSO includes a large number of fine-grained research topics, and it can be easily updated by running Klink-2 on recent corpora of publications.

The current version of CSO includes 13K semantic topics and 133K relationships. The main root is Computer Science; however, the ontology includes also a few secondary roots, such as Linguistics, Geometry, Semantics, and so on.

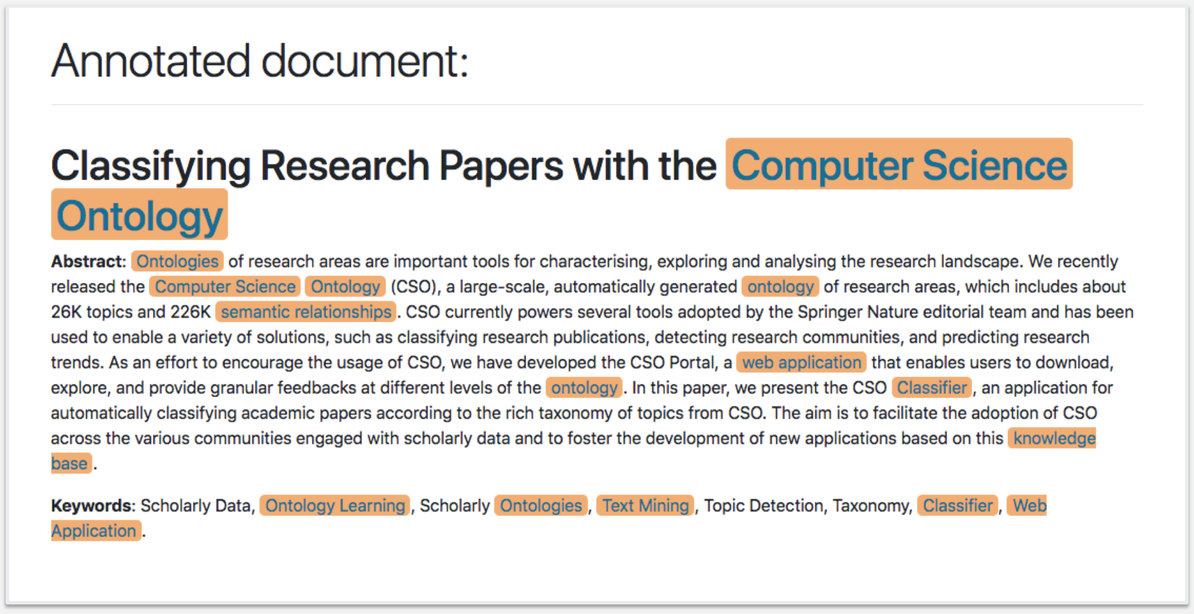
## The CSO Classifier

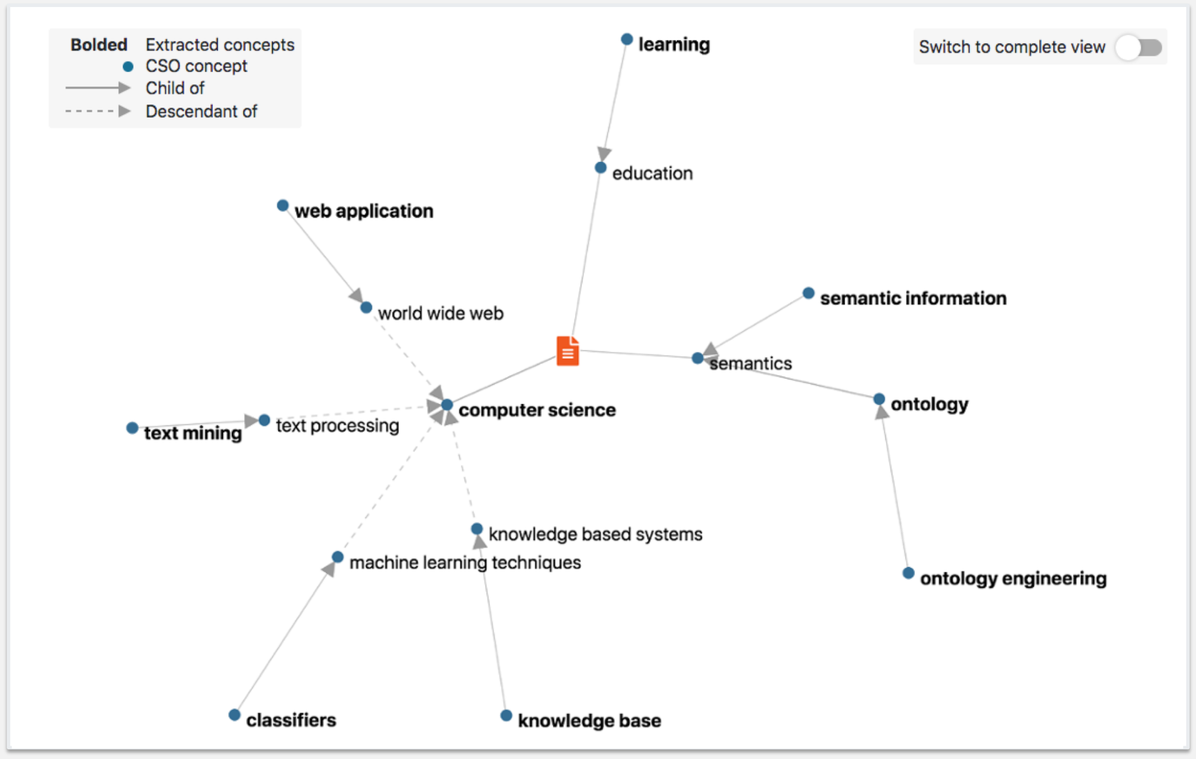
The CSO Classifier is an application that classifies the content of scientific papers (i.e., full-text, abstract, and title) according to CSO. Given a document, the CSO Classifier initially identifies its research concepts. Then, using the *skos:broaderGeneric* relationships within the CSO, the classifier infers also their super concepts – only their direct super-topics or the complete set of super-topics.




Try it!




## Learn more

*[1] Osborne, F. and Motta, E. (2015) Klink-2: Integrating Multiple Web Sources to Generate Semantic Topic Networks, ISWC2015, USA*

*[2] Salatino, A.A., Thanapalasingam, T., Mannocci, A., Osborne, F., Motta, E. (2018) The Computer Science Ontology : A Large-Scale Taxonomy of Research Areas, ISWC 2018, USA.*

http://skm.kmi.open.ac.uk/

angelo.salatino@open.ac.uk

Knowledge Media Institute

The Open University