

## CSO Documentation (1.0)

The Computer Science Ontology (CSO) is a large scale ontology of research areas that was automatically generated using the Klink-2 algorithm [1] on the Rexplore dataset [2], which consists of about 16 million publications, mainly in the field of Computer Science. The Klink-2 algorithm combines semantic technologies, machine learning, and knowledge from external sources to automatically generate a fully populated ontology of research areas. Some relationships were also revised manually by experts during the preparation of two ontology-assisted surveys in the field of Semantic Web and Software Architecture. The main root of CSO is Computer Science, however the ontology includes also a few secondary roots, such as Linguistics, Geometry, Semantics, and so on.

CSO presents two main advantages over manually crafted categorisations used in Computer Science (e.g., 2012 ACM Classification, Microsoft Academic Search Classification). First, it can characterise higher-level research areas by means of hundreds of sub-topics and related terms, which enables to map very specific terms to higher-level research areas. Secondly, it can be easily updated by running Klink-2 on a set of new publications. A more comprehensive discussion of the advantages of adopting an automatically generated ontology in the scholarly domain can be found in [3].

We are currently working on a paper fully dedicated to CSO. In the meantime, please refer to it by using the paper:

*Osborne, F. and Motta, E. (2015) Klink-2: Integrating Multiple Web Sources to Generate Semantic Topic Networks, International Semantic Web Conference 2015, Bethlehem, Pennsylvania, USA*

### Data Model

The CSO model<sup>1</sup> is an extension of the BIBO ontology<sup>2</sup> which in turn builds on SKOS<sup>3</sup>. It includes five semantic relations:

- *relatedEquivalent*, which indicates that two topics can be treated as equivalent for the purpose of exploring research data (e.g., Ontology Matching, Ontology Mapping).
- *skos:broaderGeneric*, which indicates that a topic is a sub-area of another one (e.g., Linked Data, Semantic Web)
- *contributesTo*, which indicates that the research outputs of one topic contributes to another. For instance, research in Ontology Engineering contributes to the Semantic Web, but arguably Ontology Engineering is not a sub-area of the Semantic Web – that is, there is plenty of research in Ontology Engineering
- *rdf:type*, this relation is used to state that a resource is an instance of a class. For example, a resource in our ontology is an instance of topic.
- *rdfs:label*, this relation is used to provide a human-readable version of a resource's name.

### CSO Portal

We are currently building a portal for browsing, downloading, and offering feedback on CSO. Please stay tuned for it.

### Applications

CSO ontology is used by a variety of applications and methodologies:

---

<sup>1</sup> <http://technologies.kmi.open.ac.uk/rexplorer/ontologies/BiboExtension.owl>

<sup>2</sup> <http://purl.org/ontology/bibo/>

<sup>3</sup> <https://www.w3.org/2004/02/skos/>

**Smart Topic Miner.** The Smart Topic Miner (STM) [4] is a tool which uses semantic web technologies to classify scholarly publications on the basis of a very large automatically generated ontology of research areas. It was developed to support the Springer Nature Computer Science editorial team in classifying proceedings. A demo of the system is available at [http://rexplore.kmi.open.ac.uk/STM\\_demo](http://rexplore.kmi.open.ac.uk/STM_demo).

**Smart Book Recommender.** The Smart Book Recommender (SBR) [5] is a semantic application designed to support the Springer Nature editorial team in promoting their publications at Computer Science venues. It takes as input the proceedings of a conference and suggests books, journals, and other conference proceedings which are likely to be relevant to the attendees of the conference in question. A demo of the system is available at [http://rexplore.kmi.open.ac.uk/SBR\\_demo/](http://rexplore.kmi.open.ac.uk/SBR_demo/).

**Rexplore.** Rexplore [2] is a system which leverages novel solutions in large-scale data mining, semantic technologies and visual analytics, to provide an innovative environment for exploring and making sense of scholarly data.

**EDAM methodology.** EDAM [6] is a novel expert-driven automatic methodology for creating Systematic Reviews that keep human experts in the loop, but does not require them to check all papers included in the analysis.

If you are using CSO in your system, please contact us and we will add it to this list.

## Versions

Two version of CSO are currently available:

- CSO 1.0. Generated by applying Klink-2 v. 1.0 on the Rexplore dataset. It includes about 15k topics linked by 96k semantic relationships.
- CSO 2.0. Generated by applying Klink-2 v. 2.0 on the Rexplore dataset. It includes about 26k topics which are linked by 226k semantic relationships.

## Licence



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

## References

1. Osborne, F. and Motta, E. (2015) Klink-2: Integrating Multiple Web Sources to Generate Semantic Topic Networks, International Semantic Web Conference 2015, Bethlehem, Pennsylvania, USA
2. Osborne, F., Motta, E. and Mulholland, P. (2013) Exploring Scholarly Data with Rexplore, International Semantic Web Conference, Sydney, Australia.
3. Osborne, F. and Motta, E. (2012) Mining Semantic Relations between Research Areas, International Semantic Web Conference, Boston, MA
4. Osborne, F., Salatino, A., Birukou, A., Motta, E. (2016) Automatic Classification of Springer Nature Proceedings with Smart Topic Miner. International Semantic Web Conference 2016, Kobe, Japan.

5. Osborne, F., Thanapalasingam, T., Birukou, A. and Motta, E. (2017) Smart Book Recommender: A Semantic Recommendation Engine for Editorial Products, Poster at International Semantic Web Conference 2017
6. Osborne, F., Lago, P., Muccini, H., Motta, E. (2018) Reducing the Effort for Systematic Reviews in Software Engineering. In preparation.