



The SCAR Project

Semantic Coloring of Academic References

Angelo Di Iorio¹, Freddy Limpens¹, Agata Rotondi¹, Pascal Coupet², Georgios Tsatsaronis²

¹DASPLab - Digital and Semantic Publishing Laboratory (University of Bologna), ²Elsevier B.V. (Amsterdam)

<http://dasplab.cs.unibo.it/index.php/scar/>



Goal

Enrich bibliographies with metadata about individual bibliographic reference and identify authors motivations for citations. Each reference is treated as an individual, first-class entity, which can be accessed, filtered and grouped with other references according to different criteria ('coloring schemes')

Background

Outgoing Citations

References = Cited Entities

- Bobrow, D.G. and Terry, W., 1977, An Overview of KRL, a Knowledge Representation Language, Cognitive Science (1);
- Hayes, P.J., 1979, The Logic of Frames, in D. Metzging, pp 46-61;
- Minsky, M., 1975, A Framework for Representing Knowledge, in The Psychology of Computer Vision, pp 211-277;
- Schmolze, J.G. and Brachman, R.J., 1982, Summary of the KRLONE Language, in Proc of 1981 KRLONE Workshop
- [...]

Citation Context

The KRLONE representation language (Schmolze and Brachman, 1982) has a variety of notational devices, each with an associated efficient deduction procedure. Hayes (1979) has argued that frame representations (Minsky, 1975; Bobrow and Winograd, 1977) should be viewed as [...]

In-text Reference Pointers

Which Citation Functions?

Incoming Citations

Citing Entities

Citation Context

HLF (Hobbs, 1985) is another common choice of semantic representation [...]

In-text Reference Pointer

Which Citation Functions?

Methods & Technologies

- Machine Learning
 - Supervised Approach
- Semantic Web
 - RDF, SPARQL, SPAR
- NLP
 - Tokenization, POS Tagging and Parsing, Corpora Analysis, Words Frequency and Similarity

Project Timeline

Analysis of Existing Citation Functions Schemas

Analysis of Criteria for Characterizing Citations

Build RDF Dataset on Test Articles

Build an Annotated Corpus + Guidelines for Annotation

Users Evaluation

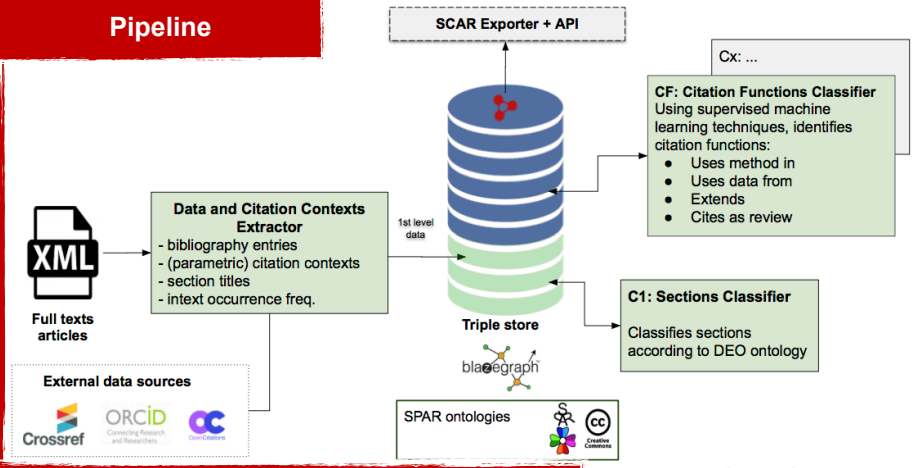
Development of a Proper Schema, Verified via Users Survey

Development of the Citation Extraction Tool: Citation Network and Textual Citation Context

Machine Learning Classifier

Validation

Pipeline

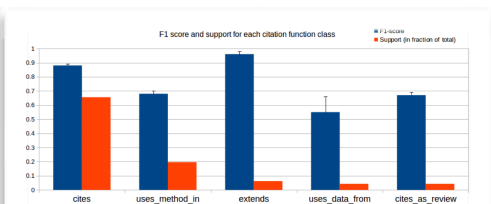


SCAR Citation Functions Schema

Uses Method In	The work of the citing entity uses a method presented in the cited entity <i>Amber force field [34] was used to model the DNA</i>
Extends	The citing entity extends, improves or continues the work of the cited entity <i>Our web application implements a Parallel Exploration paradigm which extends Parallel Faceted Browsing [9]</i>
Uses Data From	The citing entity uses data presented in the cited entity <i>We mainly experiment on the Semantic 3D dataset [40]</i>
Cites as Review	The citing entity cites a work that reviews a series of works <i>Following previous studies (see Lin's review [4]), we measured social capital [...]</i>

Validation

Result of 3-fold cross validation using ground-truth dataset of the citation functions classifier.



Annotated Corpus

- Annotated corpus about citation functions used to improve the performance of the classifier - 820 annotations (in progress)
- Ad-hoc guidelines in order to help annotators
- The annotated corpus will be soon released and freely available



scar@cs.unibo.it