A review of ontologies for describing scholarly and scientific documents

Almudena Ruiz-Iniesta and Oscar Corcho

Ontology Engineering Group
Universidad Politécnica de Madrid, Madrid, Spain
{almudenari, ocorcho}@fi.upm.es

COMENTARIO: creo que los ejemplos que voy intercalando en el texto se pueden quitar, no sé si aportan mucho a la revisión

Abstract. Semantic publishing of scholarly publications and scientific documents is a growing tend in the last years. This has led to an increasing number of ontologies for describing documents. This paper presents a classification of these state-of-the-art ontologies, and describes them. The rationale behind the different approaches is presented. Some challenges that the approaches described could face are also pointed out. Finally, we propose the use of some of this ontologies within the annotation of scientific documents.

Keywords: Semantic Web, ontologies, document semantics, semantic publishing

1 Introduction

In the last years the way in scholarly articles are published is changing due to the evolution driven by the Semantic Publishing community [1,2]. Semantic publishing enhances the publication of a document such as a journal article so as to enrich its meaning, provides a way to understand the meaning of the published information and enables the linking to related articles. The goal of semantic publication is to offer a fully open-access to both the content of the document and the metadata, that let to understand the internal structure of the documents and their links with other documents. Semantic publishing is a hot research topic and an example is the amount of different ontologies related to describe different scholarly domains, EXPO[3] in the scientific experiments domain, OMDoc a markup format and data model for Open Mathematical Docments[4], and the SWAN ontology for modeling scientific discourse developed in the context of building a series of applications for biomedical researchers [5],etc. Nevertheless, it is also important to describe the entire documents, their structure, their rhetorical elements and all of the related information that can enhance the semantic significance of the document.

We are currently studying how existing formal models for describing scientific documents. In particular, we are interested in the existing ontologies for describing scholarly documents. However, the variety of works that describe documents in different domains makes difficult to choose the better ontology that fits to our preferences and/or goals. In this paper we present a review of the existing ontologies for describing scholarly publications and we also present other vocabularies that allow to embed formal metadata in documents using markup languages. Our goal is to shed light on this variety of works in order to facilitate the choice of what it's the best ontology in each case.

The result is a classification of the most important ontologies for describing documents. The proposed classification divides the ontologies into three main groups: ontologies for describing the document structure (sections, paragraphs, etc.), ontologies for describing the rhetorical elements (introduction, results, etc.) and ontologies for describing bibliographies and cites. In what follows, we expand on these ontologies, and show how they can be employed to describe a paper. To better understand what different annotation approaches consist in, among the paper we illustrate how some of them could be applied to a published article from the journal *Future Generation Computer Systems*, that is encapsulated as a Research Object[6][1].

The rest of the paper is organized as follows: Section 2 describes the main ontologies that describes documents; Section 3 presents the works that attempts to annotate the references of a document; Section 4 introduces other vocabularies that could enhanced the annotation of a document. Finally, Section 5 concludes the paper and depicts some recommendations to annotate a scientific document.

2 Ontologies for describing documents

In this section we describe the present and past ontologies that are focused in describe the structure of a scholarly article or, more generally, of a document. Each ontology is presented with it main characteristics and an example of use.

One of the first works was Document ontology[2]. This ontology models several kinds of documents but it only defines the type of the document. Some of the documents types that we can found in this ontology are: Abstract, Letter, Form, Lecture, etc. This ontology, which was developed in the pre-Semantic Web era, as part of the SHOE project[7] a small extension to HTML to allow web page authors to annotate their web documents with machine-readable knowledge, currently this work is not maintained. The following bullet shows how to use Document ontology approach to annotate the journal article, we would include information like the following:

- An instance *Future-Generation-Computer-Systems* of the `Journal` category.
- `author(Document, Person)` relationship that indicates who is the author. Person is another category from the general ontology. In this case the relationship will established among *Paper1-FGCS* and *Daniel-Garijo* (instance from Person).

[1] `http://rohub.linkeddata.es/motifs_bundle_page-FGCS/`

[2] `http://www.cs.umd.edu/projects/plus/SHOE/onts/docmnt1.0.html`

- `title(Document, .STRING)` relationship that indicates the title of the publication. title(Paper1-FGCS, “Common Motifs in Scientific Workflows: An Empirical Analysis”).
- `volume(Periodical, .NUMBER)` relationship that describes information about the publication.volume(*Future-Generation-Computer-Systems*, *In press*).

Ontology of Rhetorical Blocks (ORB)[3] is a formalization capturing the coarse-grained rhetorical structure of scientific publications independently of the domain. The ontology models a publication by means of three artefacts: the header, the body and the tail. The header is an ontological entity itself, is the part of the publication that models meta-information about the publication, including fields such as title, authors, affiliations, publishing venue or abstract. The body is composed by four rhetorical blocks: introduction, methods, results and discussion according to the IMRAD[8] structure. Each one of these blocks are refereed to an ontological entity. Finally the tail provides additional meta-information about the paper related to external references. The tail is represented by two ontological entities: acknowledgments and references. As we mentioned above this ontology provides a high-level approach for describing documents. ORB can be used to define content in RDF format[9].

The following snippet is showing how to use ORB to annotate the journal article, it combines ORB and Dublin Core[10][4].

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix orb: <http://purl.org/orb/>

[    dc:creator "Garijo,␣Daniel" ;
   dc:creator "Alper,␣Pinar" ;
   dc:creator "Belhajjame,␣Khalid" ;
   dc:creator "Corcho,␣Oscar" ;
   dc:creator "Gil,␣Yolanda" ;
   dc:creator "Goble,␣Carole" ;
     dc:title "Common␣motifs␣in␣scientific␣workflows..." ;
     dcterms:abstract "Workflow␣technology␣continues␣to␣play
         ..."
     dcterms:hasPart
         [ a orb:Header ;
             dc:description "Header␣content"
         ],
         [ a orb:Introduction ;
             dc:description "Introduction␣content"
         ]
         ...
         [ a orb:References ;
           dc:description "List␣of␣references"
```

[3] http://www.w3.org/2001/sw/hcls/notes/orb/#ontology

[4] This and the following RDF examples are written in Turtle (http://www.w3.org/TeamSubmission/turtle/)

```
          ]
].
```

Finally we can find Semantic Publishing and Referencing Ontologies (SPAR, namespace `http://purl.org/spar`) a set of several ontologies that provide the ability to describe books and journal articles, by enabling RDF metadata to be created to relate these entities to reference citations, to bibliographic records, to the component parts of documents, and to various aspects of the scholarly publication process. This set of ontologies is composed by:

- FaBiO[5], the FRBR-aligned Bibliographic Ontology, is an ontology for recording and publishing on the Semantic Web bibliographic records of scholarly endeavours.
- CiTO[6][11], the Citation Typing Ontology, is an ontology for the characterization of citations, both factually and rhetorically.
- BiRO[7], the Bibliographic Reference Ontology is an ontology for describing bibliographic records and references, and their compilation into bibliographic collections and reference lists.
- C4O[8], the Citation Counting and Context Characterization Ontology allows the characterization of bibliographic citations in terms of their number and their context.
- DoCO[9], Document Components Ontology is an ontology for describing the components parts of a document. DoCO imports the Discourse Elements Ontology[10] and the Document Structural Patterns Ontology[11].
- PRO[12], the Publishing Roles Ontology is an ontology for the characterization of the roles of agents in the publication process.
- PSO[13], the Publishing Status Ontology is an ontology for characterizing the publication status of a document at each of the various stages in the publishing process.
- PWO[14], the Publishing Workflow Ontology is an ontology for describing the steps in the workflow associated with the publication of a document.

As we can see this all set of ontologies let us to semantically-enhanced the entire characteristics of a document, when we take together, these provide a complete semantic description of a document. In this work we are analyzing those that are focused on describing the document content and so we will describe in detail DoCO (see Section 2.1) and CiTO (see Section 3).

[5] Namespace `http://purl.org/spar/fabio/`

[6] Namespace `http://purl.org/spar/cito`

[7] Namespace `http://purl.org/spar/biro`

[8] Namespace `http://purl.org/spar/c4o`

[9] Namespace `http://purl.org/spar/doco`

[10] Namespace, `http://purl.org/spar/deo`

[11] Namespace, `http://www.essepuntato.it/2008/12/pattern`

[12] Namespace `http://purl.org/spar/pro`

[13] Namespace `http://purl.org/spar/pso`

[14] Namespace `http://purl.org/spar/pwo`

2.1 DoCO, Documents Components Ontology

COMENTARIO: Dijimos de dotar a DoCO de entidad propia, pero quizá no sea necesario...

In order to describe a document based on its structure and content the DoCO ontology provides a broad number of classes and relationships that allow this. DoCO import two ontologies, Deo and the Document Structural Patterns Ontology. Deo is an ontology written in OWL 2 DL for describing the major rhetorical elements of a document such as a journal article. It also provides a structured vocabulary for rhetorical elements within documents and it uses all the rhetorical block elements from the SALT Rhetorical Ontology [12]. On the other hand, the pattern ontology defines formally patterns for segmenting a document into atomic components, in order to be manipulated independently and re-flowed in different contexts.

DoCO describes the vast majority of document components such as chapter, preface, glossary, etc. Our example article described with DoCO looks like shows the next snippet.

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix doco : <http://purl.org/spar/doco/> .

<http://dx.doi.org/10.1016/j.future.2013.09.018> #doi of the
    journal article
  doco:Paragraph ;
  doco:Section ;
].
```

As far as Deo is concerned, Deo is a versatile ontology that supports the main rhetorical elements in a document. These rhetorical components, for example Introduction, Methods, Results and Conclusions, give a defined rhetorical structure to the paper, which assists readers to identify the important aspects of the paper. Notice that the rhetoric organization of a paper does not necessarily correspond neatly to its structural components (sections, paragraphs, etc.). In this sense, Deo and DoCO complement one another. Table 1 shows some of the classes from Deo.

Table 1. List of some of Deo classes

acknowledgements	background	conclusion
introduction	future work	methods
related work	results	discussion
motivation	problem statement	biography

2.2 Scholarly and Scientific discourse ontologies

Scientific discourse has particular characteristics that are not covered at all by the above ontologies. Particularly, a scientific discourse has goals, claims, experiments, evaluations and so on. Indeed, the reasoning of the assertion of the scientific document is crucial for scholarly and scientific publishing, in proposing hypotheses and advancing evidence in their support. Several works have been proposed to model the discourse argumentation scientific articles. For instance [13,14] identifies the main components of scientific investigations and construct CISP metadata about the content of papers. The main classes proposed in CISP are: Goal of investigation, Motivation, Object of investigation, Research method, Experiment, Observation, Result and Conclusion. CISP metadata is constructed with an ontology methodology and makes use of an ontology of experiments EXPO[3] as a core ontology. EXPO[15] is a very complete ontology about scientific experiments. The aim of this ontology is to provide a controlled vocabulary of scientific experiments. For this purpose EXPO defines over 200 concepts for creating semantic markup about scientific experiments. This aims to provide a formal description of experiments for efficient analysis, annotation and sharing of results. EXPO is able to describe computational and physical experiments, experiments with explicit and implicit hypothesis. EXPO defines general classes including ScientificExperiment, ExperimentGoal, ExperimentTechnology, ExperimentResult, etc.

Others works use the 'Toulmin Model of Argument'[15]. That is the case of The Argument Model Ontology[16] an OWL 2 DL ontology that allows to describe argumentation according to the Toulmin's theory. This ontology aims to encode the Toulmin's theory through OWL classes and properties, in order to describe a web of inter-linked entities that participate, with a specific role, in one or more arguments. Moreover, this ontology is aligned with CiTO, the Citation Typing Ontology (`http://purl.org/spar/cito`), an ontology for the characterization of citations, both factually and rhetorically (CiTO will be described in Section 3). Toulmin proposed a layout containing six interrelated components for analyzing arguments: claim, evidence, warrant, backing, qualifier and rebuttal. The Argument Model Ontology models this components through a set of 8 classes and 21 properties. The following snippet shows an example of use of this ontology:

```
: sentence1 dcterms : description "␣We␣propose␣a␣catalog␣of␣
    domain␣independent␣conceptual␣abstractions␣for␣workflow␣
    steps␣that␣we␣call␣scientific␣workflow␣motifs␣" .
: sentence2 dcterms : description "␣We␣present␣an␣empirical␣
    analysis␣performed␣over␣260␣scientific␣workflow␣
    descriptions.␣␣" .
: sentence3 dcterms : description "This␣paper␣extends␣our␣
    previous␣work,␣which␣performed␣an␣analysis␣of␣177␣
    workflows" .
```

[15] (http://expo.sourceforge.net/)

[16] http://www.essepuntato.it/2011/02/argumentmodel

```
: argument1 a amo: Argument ;
  amo: hasClaim : sentence1 ;
  amo: hasEvidence : sentence2 ;
  amo: hasWarrant : sentence3 .
: argument2 a amo: Argument ;
  amo: hasClaim : sentence3 ; # etc .
```

Beyond the works that employ a linguistic model exists other works that are focused in describe the scientific discourse itself and the relations among the claims and hypothesis made by the author of the document. That is the case of the last two works that we present here.

The SWAN[17] ontology [5] is an ontology for modeling scientific discourse. The aim of the SWAN ontology is to enable a social-technical ecosystem in which semantic context of scientific discourse can be created, stored, accessed, integrated and exchanged along with unstructured or semi-structured digital scientific information. The SWAN project is part of the Annotation Ontology[16] and it has evolved into Domeo annotation toolkit[18]. Domeo is an extensible web application enabling users to visually and efficiently create and share ontology-based stand-off annotation on HTML or XML document targets.

The core of the SWAN ontology models the discourse elements providing a model of assertions, questions and hypothesis. The SWAN discourse elements are:

– Research statements: a claim or an hypothesis.
– Research questions: topics under investigation.
– Structured comments: the structured representation of a comment published in a digital resource.

The last work that focus on model the scholarly discourse is ScholOnto[17]. The ScholOnto ontology provides a small set of uncontroversial conceptual and relational types which are simple enough to understand without being simplistic, yet expressive enough that most researchers can express the key claims made in most documents. The main class of the ontology is the *Claim*. All claims are owned by an agent, and have some form of justification. Claims assert new relationships with other claims, or between concepts. Figure 1 shows the structure of a scholarly *Claim* in the ScholOnto ontology.

3 Ontologies for describing bibliography and cites

The Citation Typing Ontology (CiTO, namespace `http://purl.org/spar/cito`) is an ontology to enable characterization of the nature or type of citations, both factually and rhetorically. CiTO provides a set of object properties that adds more information to the cite (e.g.agress with, corrects, likes, uses method in, etc.). CiTO allows to characterize the citations in three ways: explicit citations

[17] Semantic Web Applications in Neuromedicine http://www.w3.org/TR/hcls-swan/
[18] http://swan.mindinformatics.org/

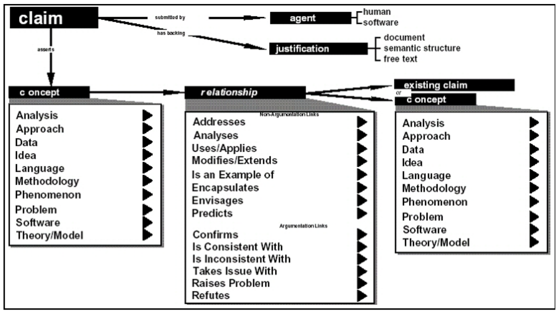

Fig. 1. The structure of a scholarly *Claim* in the ScholOnto ontology

(e.g. the reference list of a journal article), indirect citations (e.g. a citation to a more recent paper by the same research group on the same topic), or implicit citations (e.g. as in artistic quotations or parodies, or in cases of plagiarism).

Some of the CiTO properties are summarized in Table 2

Table 2. List of some of CiTO properties

agrees with	cites as authority	cites as evidence	confirms	corrects
describes	disagrees with	extends	includes excerpt from	discusses
supports	updates	uses conclusions from	uses data from	uses method in

Let us see an example of using CiTO for characterizing the citations of the journal paper.

```
@prefix cito: <http://purl.org/spar/cito/> .

:paperA cito:agrees with :paperB ;
        cito:supports :paperC ;
```

The other widely adopted ontology for characterizing bibliographies is Bibliographic Ontology Specification (BIBO, namespace `http://purl.org/ontology/bibo/`). BIBO defines a set of classes to identify the type of document based on its origin (journal, book, webpage, etc.). BIBO is implemented in RDF. This ontology could be useful to classify the documents according his provenance.

Let us suppose that we want to annotate the journal paper proposed with BIBO. In this case we would include information like the following:

```
@prefix bibo: <http://purl.org/ontology/bibo> .
@prefix dc: <http://purl.org/dc/terms> .
```

```
[ a bibo:Article ;
  dc:title "Common␣motifs␣in␣scientific␣workflows..." ;
 ...
]
```

4 Other vocabularies for describing documents

After we have described controlled vocabularies and formal ontologies for describing scholarly and scientific documents now, we are showing how to use other vocabularies in order to annotate the documents with more information about themselves such as authors, affiliations, keywords, journal, year, etc. For this purpose we can found several vocabularies that lets to identify other actors in the document.

Dublin Core Metadata Terms (DCT, namespace `http://dublincore.org/documents/dcmi-type-vocabulary/`) [18] is a vocabulary of fifteen properties for use in specifying the characteristics of electronic documents. The terms in DCT are intended to be used in combination with terms from other vocabularies, as we saw above in the ORB vocabulary, DCT can be exploited to add more information to the resource such as author, audience (students, lectures, etc.). It specifies a predefined set of document features such as *creator, date, contributor, description, format*, etc. DCT annotations can be implemented in languages like HTML or XHTML, RDF/XML and plain XML. We believe that DCT is not useful for expressing citations or document content but it could be useful to express the characteristics of the resource itself such as title, date, format and license.

```
@prefix dc: <http://purl.org/dc/elements/1.1/>
@prefix dcterms: <http://purl.org/dc/terms/>
@prefix dctype: <http://purl.org/dc/dcmitype/>

<http....> # url of the journal article
 dc:title "Common␣motifs␣in␣scientific␣workflows..." ;
 dcterms:dateSubmitted "2013-02-01"^^dcterms:W3CDTF  ;
 dcterms:dateAccepted "2013-09-05"^^dcterms:W3CDTF ;
 dcterms:format mime:application/pdf ;
 dcterms:license      <http://creativecommons.org/licenses/
    by/3.0/> ;
```

Friend of a Friend (FOAF)[19] aims at linking people and information using the Web. FOAF has been evolving gradually since its creation in mid-2000, currently there is a stable ontology that contains some classes such as Agent, Person, Organization, Group, Project, Document, Image, etc., and some properties to describe the instances of these classes. This ontology is implemented in RDF Schema. This vocabulary lets to describe the authors of a documents, her affiliations and other relevant information about it.

[19] `http://www.foaf-project.org/`

Let us see how to apply FOAF to annotate the journal article proposed as example.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<#DG>
    a foaf:Person ;
    foaf:name "Daniel␣Garijo" ;
    foaf:mbox <mailto:dgarijo@fi.upm.es> ;
    foaf:knows [
        a foaf:Person ;
        foaf:name "Oscar␣Corcho"
    ] .
```

5 Conclusions and recommendations

In this paper we have described the most ontologies used for describing documents together with an example of use for each one. Moreover, we have also sketched out a model for the enhancements of documents based on some of this ontologies, mainly those that describe the document structure, the rhetorical elements and the references. We have classified the ontologies in two main groups, those that describe the discourse of a document either scientific or generic, and those that allow to describe the document structure. In addition have been presented those ontologies dedicated to describe references and citations. At last, we have presented others well-known vocabularies which allows to provide more information about a document.

(**TODO TODO TODO:** Finalmente mostramos un conjunto de recomendaciones sobre cómo utilizar el conjunto de todas estas ontologías para describir una artículo completo, y cuál es nuestra aproximación para describir artículos científicos. Firstly we propose to use the DoCO ontology for describing the document structure. In Figure XX we can see how the Title, header, section and so on classes are employed. To describe an article in this way allow to compare two or more papers according its structure and it could be possible to detect PLAGIOS. otro posible uso, El tener la estructura de muchos artículos puede servir para generar plantillas que sirvan a jóvenes investigadores a iniciar su carrera. Después de tener descrita la estructura del documento, pronemos utilizar Deo para anotar la retórica del discurso, hay que señalar que para cada dominio concreto de aplicación la retórica debe ser ampliada y más concreta para el discurso científico. Si bien es cierto que las clases que Deo proporciona son lo suficiente genéricas como para poder ser aplicada en cualquier dominio siempre es necesario adaptarla un poco. En el ejemplo que presentamos en la Figura XX hemos extendido DEO con una clase propia para describir el discurso científico como es Hypothesis. Hemos definido esta clase de una manera formal como, an educated guess about things work. A hypothesis should be something that you can test. Para la parte de referencias y citas hemos utilizado las dos ontologías

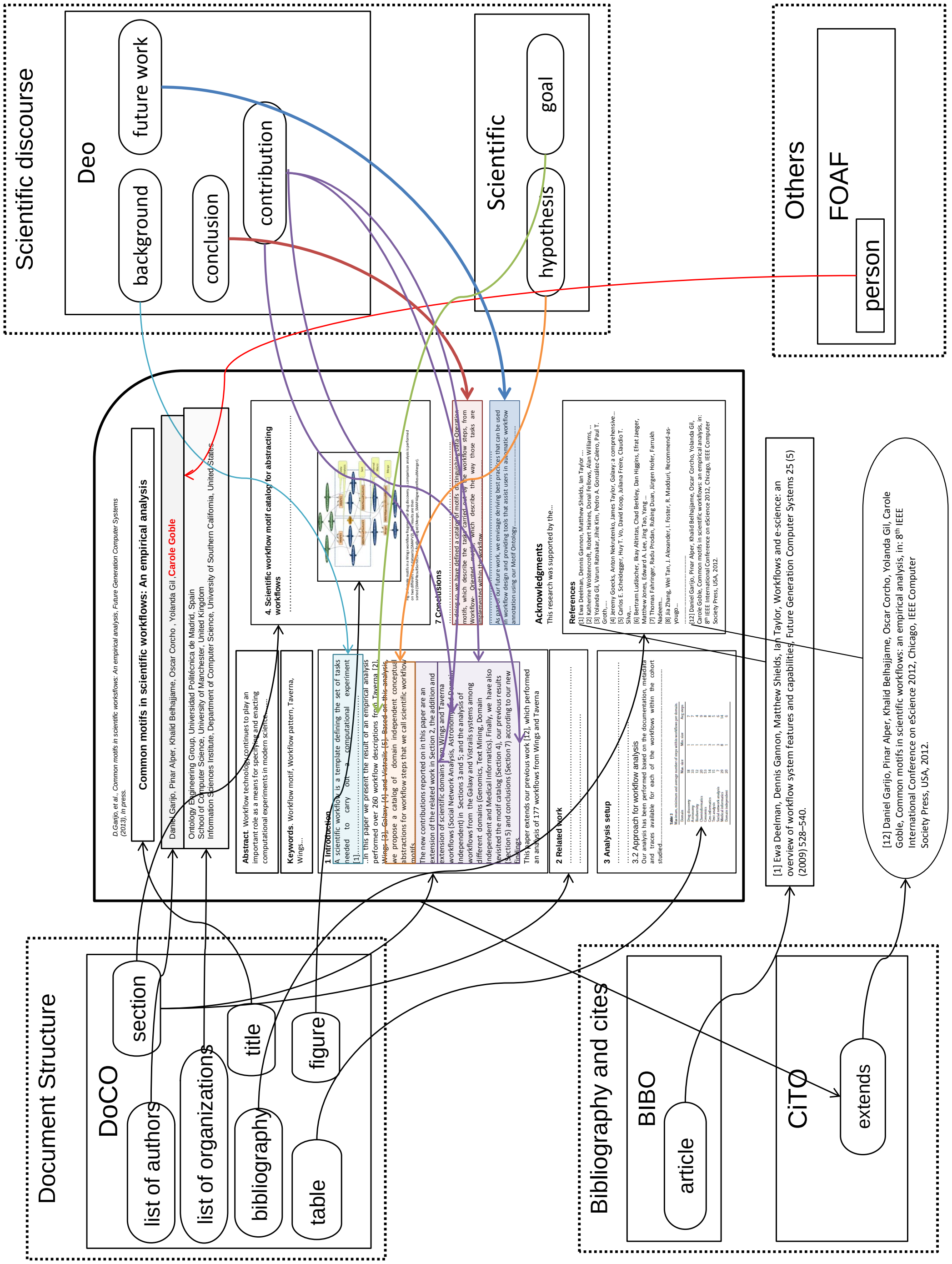
Scientific discourse
Deo
background
conclusion
contribution
future work
Scientific
hypothesis
goal
Others
FOAF
person
Common motifs in scientific workflows: An empirical analysis
Daniel Garijo, Pinar Alper, Khalid Belhajjame, Oscar Corcho , Yolanda Gil, Carole Goble
Ontology Engineering Group, Universidad Politécnica de Madrid, Spain
School of Computer Science, University of Manchester, United Kingdom
Information Sciences Institute, Department of Computer Science, University of Southern California, United States
Abstract
Keywords
1 Introduction
2 Related work
3 Analysis setup
4. Scientific workflow motif catalog for abstracting workflows
7 Conclusions
Acknowledgments
References
[1] Ewa Deelman, Dennis Gannon, Matthew Shields, Ian Taylor, Workflows and e-science: an overview of workflow system features and capabilities, Future Generation Computer Systems 25 (5) (2009) 528–540.
[12] Daniel Garijo, Pinar Alper, Khalid Belhajjame, Oscar Corcho, Yolanda Gil, Carole Goble, Common motifs in scientific workflows: an empirical analysis, in: 8th IEEE International Conference on eScience 2012, Chicago, IEEE Computer Society Press, USA, 2012.
Document Structure
DoCO
section
list of authors
list of organizations
title
bibliography
figure
table
Bibliography and cites
BIBO
article
CiTO
extends

Fig. 2. The proposed model for describing a scientific paper taking together the described ontologies.

presentadas BIBO y CiTO. La primera de ellas nos ha permitido describir de una manera detallada cada una de las referencias utilizadas por el autor junto con la información bibliográfica del propio artículo (see Figure XX). La segunda ontología, CiTO, ha sido utilizada para describir las relaciones que establece el autor del documento con las distintas referencias utilizadas. De tal manera que podemos observar como en el artículo el autor *describes* otro trabajo y como *agrees* con otro autor.

when take together, these provide a complete semantic description of a scientific publication, its relationships with similar publications, and its role in the world of scholarship.)

Acknowledgments

This work is supported by the FP7 European project Dr Inventor FP7-611383.

References

1. Shotton, D.: Semantic publishing: the coming revolution in scientific journal publishing. Learned Publishing **22**(2) (April 2009) 85–94
2. Waard, A.d.: From proteins to fairytales: Directions in semantic publishing. IEEE Intelligent Systems **25**(2) (2010) 83–88
3. Soldatova, L.N., King, R.D.: An ontology of scientific experiments. Journal of The Royal Society Interface **3**(11) (December 2006) 795–803 PMID: 17015305.
4. Kohlhase, M.: Omdoc: Open mathematical documents. In: OMDoc An Open Markup Format for Mathematical Documents. Number 4180 in Lecture Notes in Computer Science. Springer Berlin Heidelberg (January 2006) 25–32
5. Ciccarese, P., Wu, E., Wong, G., Ocana, M., Kinoshita, J., Ruttenberg, A., Clark, T.: The SWAN biomedical discourse ontology. Journal of Biomedical Informatics **41**(5) (October 2008) 739–751
6. Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., Couch, P., Cruickshank, D., Delderfield, M., Dunlop, I., Gamble, M., Michaelides, D., Owen, S., Newman, D., Sufi, S., Goble, C.: Why linked data is not enough for scientists. Future Generation Computer Systems **29**(2) (February 2013) 599–611
7. Heflin, J., Hendler, J., Luke, S.: SHOE: a knowledge representation language for internet applications. Technical report (October 1999)
8. Sollaci, L.B., Pereira, M.G.: The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. Journal of the Medical Library Association **92**(3) (July 2004) 364–371
9. W3C: Rdf primer : W3c recommendation 10 february 2004 (2004)
10. Dublin Core Metadata Initiative: Dublin core metadata element set, version 1.1
11. Peroni, S., Shotton, D.: FaBiO and CiTO: ontologies for describing bibliographic resources and citations. Web Semantics: Science, Services and Agents on the World Wide Web **17** (December 2012) 33–43
12. Groza, T., Handschuh, S., Möller, K., Decker, S.: SALT - semantically annotated LaTeX for scientific publications. In Franconi, E., Kifer, M., May, W., eds.: The Semantic Web: Research and Applications. Number 4519 in Lecture Notes in Computer Science. Springer Berlin Heidelberg (January 2007) 518–532

13. Liakata, M., Teufel, S., Siddharthan, A., Batchelor, C.R.: Corpora for the conceptualisation and zoning of scientific papers. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D., eds.: Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta, European Language Resources Association (2010)
14. Soldatova, L., Liakata, M.: An ontology methodology and CISP - the proposed core information about scientific papers. Programme/Project deposit (December 2007)
15. Toulmin, S.E.: The Uses of Argument. Cambridge University Press (July 2003)
16. Ciccarese, P., Ocana, M., Garcia Castro, L., Das, S., Clark, T.: An open annotation ontology for science on web 3.0. Journal of Biomedical Semantics **2**(2) (2011) 1–24
17. Shum, S.B., Motta, E., Domingue, J.: ScholOnto: an ontology-based digital library server for research documents and discourse. International Journal on Digital Libraries **3**(3) (October 2000) 237–248
18. : Dublin core metadata element set, version 1.1 (2012)