Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 137 (2018) 151–162

Procedia Computer Science

www.elsevier.com/locate/procedia

SEMANTiCS 2018 – 14th International Conference on Semantic Systems

SemSur: A Core Ontology for the Semantic Representation of Research Findings

Said Fathalla[a,b,*], Sahar Vahdati[a], Sören Auer[c,d], Christoph Lange[e]

[a] *Smart Data Analytics (SDA), University of Bonn, Germany*
[b] *Faculty of Science, University of Alexandria, Egypt*
[c] *L3S Research Center, Leibniz University of Hannover, Germany*
[d] *TIB Leibniz Information Center for Science and Technology, Hannover, Germany*
[e] *Fraunhofer IAIS, Germany*

Abstract

The way how research is communicated using text publications has not changed much over the past decades. We have the vision that ultimately researchers will work on a common structured knowledge base comprising comprehensive semantic and machine-comprehensible descriptions of their research, thus making research contributions more transparent and comparable. We present the SemSur ontology for semantically capturing the information commonly found in survey and review articles. SemSur is able to represent scientific results and to publish them in a comprehensive knowledge graph, which provides an efficient overview of a research field, and to compare research findings with related works in a structured way, thus saving researchers a significant amount of time and effort. The new release of SemSur covers more domains, defines better alignment with external ontologies and rules for eliciting implicit knowledge. We discuss possible applications and present an evaluation of our approach with the retrospective, exemplary semantification of a survey. We demonstrate the utility of the SemSur ontology to answer queries about the different research contributions covered by the survey. SemSur is currently used and maintained at *OpenResearch.org*.

Peer-review under responsibility of the scientific committee of the SEMANTiCS 2018 – 14th International Conference on Semantic Systems.

Keywords: SemSur Ontology, Semantic Metadata Enrichment, SWRL rules, Scholarly Communication, Semantic Publishing,

1. Introduction

SemSur, the Semantic Survey Ontology, is an ontology for describing research problems, approaches, implementations and evaluations in a structured and comparable way. An initial version of SemSur [8] was created for our approach towards representing research findings as a knowledge graph and has now been substantially improved and expanded. Without SemSur, the preferred way of providing researchers with an overview of a field and enabling them to compare research results was via *survey articles*. However, preparing surveys and studying the individual articles requires a significant amount of time, often several months of work. The major drawback from the reader's perspective is that most survey articles are published

* Corresponding author. Tel.: +49-163-8260501.
E-mail address: fathalla@cs.uni-bonn.de

1877-0509
Peer-review under responsibility of the scientific committee of the SEMANTiCS 2018 – 14th International Conference on Semantic Systems.
10.1016/j.procs.2018.09.015

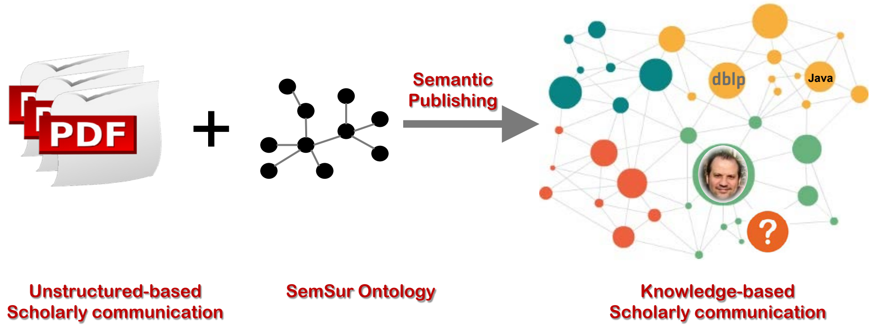

Figure 1: Transition from document-based to knowledge-based scholarly communication

in printed form or as semi-structured digital (e.g., PDF) documents, which does not make them efficiently accessible for comparative or other analyses. The SemSur ontology is a step in a long-term research agenda to create a paradigm shift from document-based to knowledge-based scholarly communication (cf. Figure 1). A striking feature of the proposed work is that it supports retrieving, exploring and comparing research findings based on an explicit semantic representation of the knowledge contained in scientific publications. If applied widely, SemSur can have a significant impact on scholarly communication. It specifically addresses researchers who want to compare their research with related works, get an overview on works in a certain field or search for research contributions addressing a particular problem or having certain characteristics. Figure 2 gives a high-level overview of SemSur by showing the five core concepts in the SemSur domain. Most prior work on semantic representation of scholarly communication focused on either describing the document structure or bibliographic information, but not individual research problems, approaches, implementations and evaluations in a machine-understandable format. In the initial version of SemSur, we identified the high level of abstraction of concepts, the lack of alignment and linking to further related ontologies, and the lack of rules for elicitation of implicit relations such as co-authorship. This article gives a comprehensive overview on SemSur, focusing on the main new features compared to the initial version, including broader coverage of the domain, better alignment with external ontologies and rules for eliciting implicit knowledge. The open W3C standards RDF and OWL were used to develop SemSur, SWRL for maintenance and quality checks as well as SPARQL for querying data adhering to SemSur. The ontology is indexed in Linked Open Vocabularies (LOV) (`http://lov.okfn.org/dataset/lov/vocabs/semsur`) and it is publicly available at `http://purl.org/semsur/owl/`, subject to the Creative Commons Attribution license. The documentation is available at `http://sda.tech/SemSur/Documentation/Semsur.html`.

The remainder of the article is structured as follows: Section 2 presents an overview on related work. Section 3 outlines the methodology for developing SemSur. Section 4 describes the of specification SemSur. An example scenario is given in Section 5. Section 6 presents an evaluation of typical usage scenarios and queries for exploring the knowledge graph in these scenarios. Section 7 concludes with an outlook on future work.

2. Related Work

This section focuses on research on ontologies describing content and structure of scholarly articles. Our review of the recent literature on scholarly communication found that most studies focus only on developing ontologies for a particular part of scholarly articles, such as article structure, bibliographic information or scientific events.

Ontologies for describing scholarly articles. Our Semantic Survey Ontology (SemSur) is a comprehensive ontology for capturing the content of survey articles involving research approaches, problems, implementations, publications and evaluations [8]. It enables building a knowledge graph for representing research findings in a structured and semantic way. The Ontology of Rhetorical Blocks (ORB) is a formalization capturing the coarse-grained rhetorical structure of scientific publications [3]. *Semantic Web for Research Communities* (SWRC) is an ontology for modeling entities of research communities such as persons, organizations, publications (bibliographic metadata) and their relationships [27]. The *Scholarly Article* (SA)[1] ontology comprises a set of concepts related to published articles such as *article*, *keywords*, *contributor* and *citation*. Moreover, it comprises a set of properties such as *roleAffiliation* and *dateRejected*. The *scientific EXPeriments Ontology* (EXPO) is a core ontology for scientific experiments, formalizing the generic concepts of experimental design, methodology, and results representation [25]. Discourse Elements Ontology (DEO) is an ontology for describing articles in terms of their main components such as Abstract, Introduction, Reference List and Figures [4]. Linked Science Core is an ontology for describing scholarly communication resources involving Publication, Researcher, Method, Hypothesis and Conclusion [2].

Ontologies for describing bibliographic citations. The *Citation Typing Ontology* (CiTO) provides a set of object properties related to citing published articles, such as "*is cited by*" and "*cites*" [19]. The *Bibliographic Ontology* (BIBO) covers the main concepts and properties for describing citations and bibliographic references [5].

Ontologies for describing scientific events. The *Semantic Web Conference Ontology* (SWC)[2] models knowledge about conferences. It covers the two sub-domains of *describing papers*, including the authors and their affiliations, and *modelling the roles* that persons can have at a conference, such as being program chair. The *Scholarly Event Description Ontology* (SEDE) is a comprehensive ontology for describing scholarly events in terms of agents (e.g., persons, committees), places (e.g., cities, venues) and time (e.g., start/end date) [15].

In this section, we have listed 12 existing ontologies describing the content and structure of scholarly articles and we have reused nine of them, as they provided suitable concepts for SemSur's domain of interest. What additionally distinguishes our work from related work is that SemSur is able to represent scientific results and to publish them in a comprehensive knowledge graph, which provides an efficient overview of a research field and compare research findings with related works in a structured way saving a significant amount of time and effort.

3. Methodology

In this section, we describe the methodology we have followed to create SemSur. The development of SemSur is driven by the following steps: exploring the domain, consulting experts, studying the literature and reusing related ontologies, defining missing concepts, defining inference rules, and finally implementing the ontology. We have iteratively interviewed several experts including ontology engineering experts and domain experts during the whole process in order to improve the final ontology. The respective stages during the ontology creation were:

- *Exploring the domain* – every research domain has its own culture, requirements and findings, e.g., technical fields such as computer science have implementations, and other fields such as agriculture have different concepts like machines, etc.). In this step, an expert in a research domain is expected to explore the culture, needs and findings of that domain and define concepts based on that. Some concepts, such as *research problem*, are needed in all research domains.
- *Asking experts* – brainstorming with other domain experts on the concepts defined from the domain to validate them: remove or update them, or add further ones, depending on the experts' feedback.
- *Reusing ontologies* – in the ontology refinement process, we explore the terms of already existing vocabularies to select the best matches and reuse or align with them.
- *Adding missing concepts* – if existing vocabularies comprise the identified concepts, then we reuse these directly, specialize them or add a property restriction, otherwise we define them in a SemSur-specific way.

[1]SA: http://ns.science.ai/
[2]SWC: http://www.scholarlydata.org/ontology/doc/

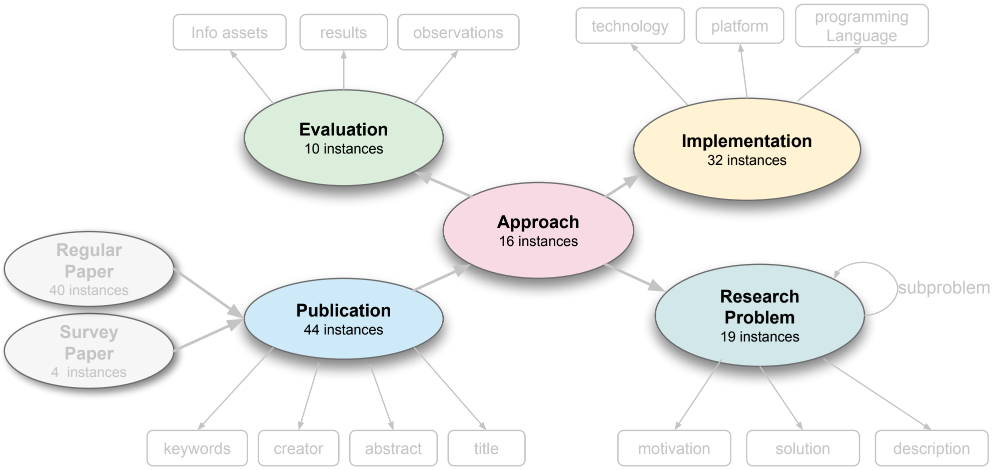

Figure 2: The five core concepts in the SemSur domain

- *Defining inference rules* – define inference rules, mostly, for properties which can be defined implicitly.
- *Implementing the ontology* – using an ontology editor and adding labels, descriptive comments and metadata.

4. SemSur Specifications

Ontology development is an iterative process that aims at producing an efficient and well-formed ontology. According to best practices (e.g., [21]), developing ontologies becomes easy and efficient by reusing existing ontologies. It is well known that when developing a semantics-based application, it is important to reuse and integrate existing ontologies to provide the background knowledge required by the application [9]. Ontology reuse should begin by identifying candidate ontologies to be reused, having them evaluated by domain experts, and choosing the most adequate ontologies to be reused among the candidate ontologies having the highest quality [21]. Pinto describes two different ontology reuse processes: fusion/merging and composition/integration [22]. *Fusion* means building ontologies by unifying knowledge from source ontologies in the same domain as the target ontology. *Composition* means building ontologies by assembling two or more ontologies that might come from domains different from the domain of the target ontology [20]. In this work we performed an ontology fusion by re-using seven ontologies from the domain of scholarly communication. We followed ontology development guidelines proposed by [22, 13], which are used by many researchers[3], to update SemSur. The ontology is structured around five core concepts:

- *Research problem* – represents a challenge in a particular field, possibly hierarchically decomposed into sub-problems, which have related problems, a motivation and possible approaches addressing the problem,
- *Approach* – comprises research methods and procedures to address a particular research problem,
- *Implementation* – describes the implementation of an approach in a concrete technical environment,
- *Evaluation* – describes how an implementation is evaluated using an evaluation method in a defined scenario,
- *Publication* – refers to an article and accompanying bibliographic metadata, including authors, title, keywords and abstract.

[3] At the time of writing, Google Scholar lists 1,500 citations for both sources, including 75 in 2017.

Table 1: Prefixes and namespace URIs of reused vocabularies.

Prefix	Vocabulary	URI
dc	Dublin Core Metadata Initiative (DCMI)	http://purl.org/dc/terms/
swrc	Semantic Web for Research Communities	http://swrc.ontoware.org/ontology#
foaf	Friend of a Friend ontology	http://xmlns.com/foaf/0.1/
mls	Machine Learning Schema	https://www.w3.org/ns/mls#
deo	The Discourse Elements Ontology	http://purl.org/spar/deo/
lsc	Linked Science Core Vocabulary	http://linkedscience.org/lsc/ns#
doap	Description of a Project	http://usefulinc.com/ns/doap#
doco	Document Components Ontology	http://purl.org/spar/doco
expo	Scientific EXPeriments Ontology	http://www.hozo.jp/owl/EXPOApr19.xml/

4.1. Alignment with and reuse of external ontologies

Since reuse increases the value of semantic data, the first step in updating SemSur was reusing vocabularies from scholarly communication-related ontologies on the Web. We have selected the most closely related ontologies listed in the Linked Open Vocabularies (LOV) directory. SemSur is aligned and linked with the following related ontologies from three categories: 1) for *describing scholarly articles*, we reused the DC, SWRC, DoCO, EXPO and FOAF ontologies, 2) for *describing the inner structure of a scientific article* independently of the field of research, we reused DEO and LSC, and 3) for *describing concepts of specific fields of research*, we reused MLS and DOAP. *DoCO*, the Document Components Ontology, provides a structured vocabulary for document components from a structural and a rhetorical perspectives. Classes describing document structure include `doco:Section` and `doco:Chapter`; classes for rhetorical components include `doco:Acknowledgements`, `doco:Abstract`, and `doco:Appendix`. *EXPO*, the scientific EXPeriments Ontology, is a core ontology to provide a formal description of scientific experiments by formalizing the generic concepts of experimental design, methodology, and results representation. EXPO is able to describe computational and physical experiments, including experiments with explicit and implicit hypothesis. It defines a general experiment class `expo:ScientificExperiment`, specific subclasses, as well as classes for the *goal*, *requirements*, *result* and *hypothesis* of an experiment. The reused vocabularies are listed in Table 1.

We created specializations of some of the imported classes from these ontologies, such as `SurveyPaper` and `RegularPaper` as specializations of the `swrc:Publication` class. Another concern is the integration of the reused ontologies, e.g., the `swrc:Publication` class was put into a relationship with `lsc:Conclusion` via the `lsc:produces` property. Because of the space limit, Table 2 only covers the main classes defined by SemSur and reused from other ontologies. Figure 3 shows the class hierarchy from general to specific classes, i.e., from `SurveyPaper` to `Publication`. As an example for inference, the `SingleAuthorPublication` class represents publications with only a single author. Individuals of this class could be inferred by the following rule[4]:

$$SingleAuthorPublication(p) \leftarrow Publication(p) \wedge creator(p, x) \wedge creator(p, y) \wedge (x = y) \quad (1)$$

In a similar way as for the classes, Table 3 lists the main relations defined by SemSur and reused from other ontologies. SemSur has two transitive relations: for representing co-authorship between researchers and for representing that a research problem has sub-problems. Co-authorship means that there is a cooperation between two or more authors in a publication. The relation `isCoAuthorOf` is a transitive relation, when considered in the restricted scope of one publication, represents the co-authorship between two `Persons`. Furthermore, it is also a symmetric relation. In addition, new transitive relations have been defined

[4]Every formula is assumed to be universally quantified over all its free variables. The equality symbol = denotes primitive logical equality, and ≠ denotes its negation.

Table 2: Main classes defined by SemSur and reused from other ontologies

Group	Classes	Source
reused	Abstract, Appendix	doco
	ScientificExperiment, ExperimentResult, ExperimentDesign, ExperimentRequirements, ExperimentHypothesis, Model, ExperimentalModel, DomainModel	expo
	Questionnaire	fabio
	Organization, Employee, Person	foaf
	ResearchProject, DevelopmentProject, ResearchTopic, Manual, Book, MasterThesis, PhDThesis	swrc
newly defined	SurveyPaper, RegularPaper, SingleAuthorPublication, InformationAsset, Complexity, PositiveAspects, Limitations, Documentation, Toolbox, Challenges, SimulationSoftware, SingleAuthorPublication, Toolbox, Benchmark, SimulationSoftware	semsur

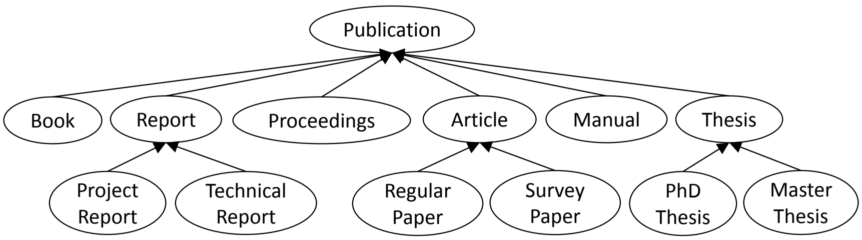

Figure 3: Publication class hierarchy of SemSur.

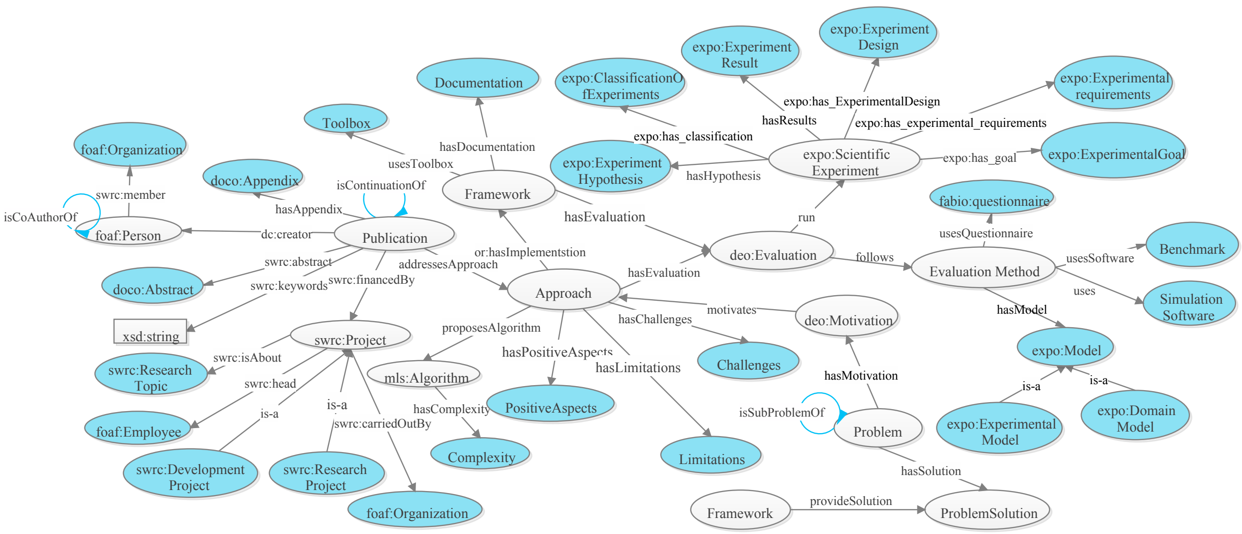

Figure 4: Overview of the SemSur 2.0 graph. New classes in SemSur 2.0 are highlighted. Some classes from SemSur 1.0 were omitted for better visualization.

such as `isContinuationOf` and `isSubProblemOf`. Also new symmetric relations have been defined, such as `isCoAuthorOf` and `hasRelatedProblem`.

Table 3: Main relations defined by SemSur and reused from other ontologies

Group	Relations	Source
reused	creator, title, hasVersion	dc
	has_experimental_requirements, has_classification, has_ExperimentalDesign, has_goal	expo
	vendor, OS, platform	doap
	name	foaf
	carriedOutBy, head, isAbout, abstract, member, financedBy, keywords	swrc
newly defined	hasLimitations, hasPositiveAspects, proposesAlgorithm, addressesApproach, isContinuationOf, hasAppendix, addressesApproach, hasChallenges, motivates, hasMotivation, provideSolution, hasSolution, hasEvaluation, hasHypothesis, hasResults, usesQuestionnaire, hasDocumentation, usesToolbox, isCoAuthorOf	semsur

4.2. SWRL Rules

A standard way to infer new information on the Semantic Web is to define inference rules [11]. Inference on the Semantic Web is used to improve the quality of data integration in the ontology by combining rules and ontologies to discover new relationships, detect possible inconsistencies and infer logical consequences from a set of asserted facts or axioms. The Protégé ontology editor can use the Drools reasoner [23] for performing rule-based inference. Several languages and standards have been proposed for writing rules for ontologies, including RuleML[5] (Rule Markup Language), Jess [10] (Java expert system shell) and SWRL (Semantic Web Rule Language) [14]. Our goal is to define a rule set for discovering new relationships and inferring new knowledge from instance data and class descriptions, which did not explicitly exist in a knowledge graph [6]. We have defined the following SWRL rules:

$$swrc : Publication(?p) \wedge dc : creator(?p, ?x) \wedge swrlb : equal(?x, 1) \rightarrow SingleAuthorPublication(?p) \tag{2}$$

$$swrc : Publication(?p) \wedge dc : creator(?p, ?x) \wedge dc : creator(?p, ?y) \wedge owl : differentFrom(?x, ?y) \rightarrow isCoAuthor(?y, ?x) \tag{3}$$

$$Problem(?x) \wedge Problem(?y) \wedge isSubProblem(?x, ?y) \wedge hasMotivation(?y, ?m) \rightarrow hasMotivation(?x, ?m) \tag{4}$$

$$swrc : Publication(?x) \wedge swrc : financedBy(?x, ?y) \wedge swrc : isAbout(?y, ?z) \rightarrow swrc : isAbout(?x, ?z) \tag{5}$$

For instance, in order to express the co-authorship between authors, we introduce the rule in Equation 3. After running Drools reasoner on the rule set, a significant number of new (ABox) axioms have been inferred. Table 4 shows the ontology statistics after running the rule engine and successful transformation of the

[5] http://wiki.ruleml.org

Table 4: A comparison between SemSur versions statistics

Metrics	SemSur 1.0	SemSur 2.0
Axioms	6,161	16,880
Logical axiom	1,696	11,260
Declaration axioms	1,076	1,867
Class	294	876
Object property	281	341
Data property	109	140
Individual	354	415
Annotation property	113	98

new knowledge into OWL. SQWRL (Semantic Query-Enhanced Web Rule Language) is a SWRL-based query language that provides SQL-like operators for extracting information from OWL ontologies [17]. The following SQWRL query is used to retrieve all single-author publications along with the author name ordered by author names.

$$SingleAuthorPublication(?p) \wedge dc:creator(?p, ?x) \rightarrow sqwrl:select(?p) \wedge sqwrl:orderBy(?x) \quad (6)$$

5. SemSur Instances: A Motivating Scenario

The final step in ontology engineering is the creation of instances/individuals of classes [1]. To better understand the domain of SemSur and to build test cases for inferencing, we have created instances representing four survey articles listed below. Currently, SemSur contains a total of 415 instances of different classes extracted from four survey articles including `Person` (126), `Publication` (4 survey papers and 40 regular papers covered by the survey papers), `implementations` (32), `approach` (16) and `ResearchProblem` (19). Table 4 shows a comparison of the different versions of SemSur in terms of ontology statistics. Figure 5 presents the core SemSur classes with one instance of each. The dashed arrow indicates the inferred statement that *Olaf Görlitz* and *Steffen Staab* are co-authors of the publication entitled "*SPARQL Endpoint Federation Exploiting VOID Descriptions*"

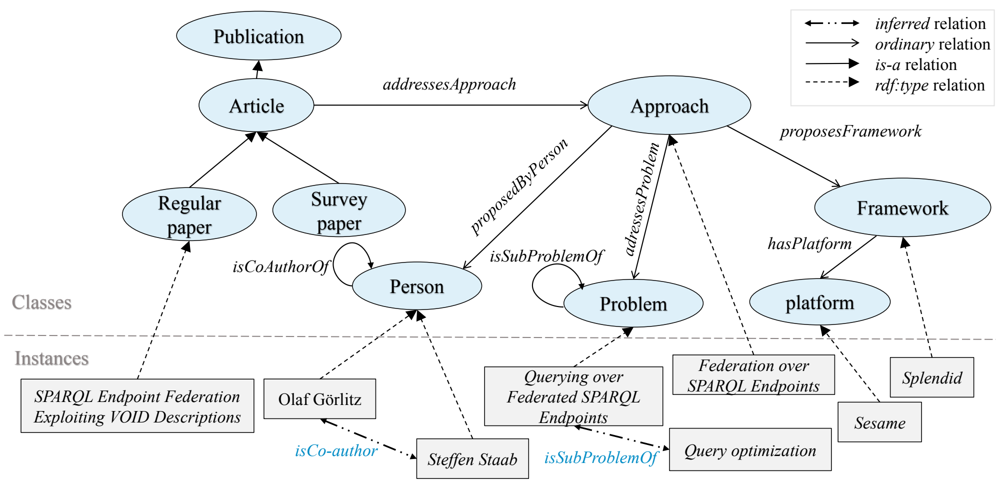

Figure 5: Instantiation of key SemSur classes

- Bringing Relational Databases into the Semantic Web: A Survey. [26]
- A Survey of Current Link Discovery Frameworks. [16]
- Querying over Federated SPARQL Endpoints –A State of the Art Survey. [24]
- Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods. [18]

We present a producer-consumer scenario in order to motivate our approach and to illustrate how SemSur can be used to complement review papers for comparative evaluation of research findings. The proposed approach is relevant for: 1) researchers who want to publish their research findings as a knowledge graph to be FAIR (findable, accessible, interoperable, reusable) for other researchers, and 2) researchers who want to get an overview of the research efforts related to a particular research problem, e.g., to write a survey paper or a related work analysis. For illustration, suppose *Alice* conducts research about “SPARQL query Federation” and publishes her research findings as a knowledge graph using the paper submission form of our OpenResearch.org platform[6]. *Bob* wants to reuse Alice’s research findings or get an overview of the state of the art of research on “SPARQL query Federation” in terms of research motivations, currently published approaches, implemented frameworks/tools, and challenges faced. *Bob* can then construct a new experiment by replacing the dataset used or modifying the approach, and then republish the new findings as a continuation of Alice’s original, including a link to her paper. Thus, *Alice* is the research producer, while *Bob* is the consumer. To support this scenario, a comprehensive ontology for describing research findings and their relationships, and a platform for adding and retrieving them are needed. Listing 1 shows the corresponding SPARQL query to achieve the goal. The output of the query includes the approach implementation, published article, research problem and motivation, and results of the experiments along with experimental requirements and goal.

Listing 1: corresponding SPARQL query to Q10 in Table 5

```
SELECT DISTINCT ?Motivation, ?Approach, ?expGoal, ?pos, ?chall
WHERE {
  or:SPARQL_query_Federation    or:hasMotivation      ?Motivation .
  ?Motivation  or:hasDescription     ?MotivationDescription .
  ?Motivation  or:motivatesApproach  ?Approach .
  ?Approach    or:hasImplementation  ?Framework .
  ?Approach    or:hasPositiveAspects ?pos .
  ?Approach    or:hasChallenges      ?chall .
  ?Approach    or:hasEvaluation      ?eval .
  ?eval        or:run                ?exp .
  ?exp         expo:has_experimental_requirements ?expGoal .
 }
```

SemSur is currently used and maintained in OpenResearch.org(OR)[7], which is a semantic crowd-sourcing platform to collect metadata about several scholarly artifacts mainly survey papers and scientific events using ontologies for each such entity type. OR is implemented on top of Semantic MediaWiki (SMW)[8] which provides a proper way of ontology implementation. Listing 2 shows an example[9] of an individual scientific paper created on OR using SemSur.

Listing 2: Description of a scientific paper and its segments on OpenResearch.org

```
{{Paper
|Title=ANAPSID:An Adaptive Query Processing Engine... |Authors=Maribel Acosta, Maria-Esther Vidal, Tomas Lampo, Julio Castillo
|Series=ISWC     |Year=2011     |Keywords=Adaptive Query Processing, ANAPSID, Linked Data  |Abstract=Following the design rules of Linked...
|Future work=In the future we plan to...      |Approach=Querying Distributed RDF Data Sources  |Problem=SPARQL Query Federation
|Implementation=ANAPSID                       |PositiveAspects=decompose the query into...
|Model=Architectural                          |Methodology=Lightweight wrappers translate SPARQL queries into...
|Challenges=Query Decomposition, Query Optimization, and Query adaptation.
|Download-page=https://github.com/anapsid/anapsid       |InfoRepresentation=RDF     |OS=Linux CentOS
|DocumentationURL=https://github.com/anapsid/anapsid    |ProgLang=Python            |Version=1.0     |GUI=No
|Subproblem=query processing on Linked Data             |RelatedProblem=decompose queries into sub-queries...
|Motivation=distribution of RDF data stores             |Benchmark=FedBench }}
```

[6] http://openresearch.org/Special:FormEdit/Paper
[7] OpenResearch.org
[8] https://www.semantic-mediawiki.org
[9] http://openresearch.org/ANAPSID:_An_Adaptive_Query_Processing_Engine_for_SPARQL_Endpoints

6. Evaluation

This section describes the evaluation procedure and then discusses the results of evaluating SemSur. We followed two strategies: a *satisfaction questionnaire* and an *expert assessment*. We divided the participants in the evaluation (all computer scientists) into two groups: experts in ontology engineering, who are aware of the challenges in this area, and researchers in other fields of computer science. To assure that we have a useful ontology, SemSur should be able to answer a number of competency queries listed in Table 5.

Satisfaction questionnaire evaluation: A total of 18 researchers, with semantic web background, were recruited for this questionnaire[10]. At the beginning of the evaluation, we made sure that all participants understood the approach by giving them: 1) a presentation about the ontology, the methodology, the domain, 2) a demonstration of a case study illustrating the potential benefits of using SemSur by trying to answer predefined queries (in Table 5) and 3) the results of a set of 20 predefined queries to measure their satisfaction of the results compared with the results of other tools (in subsection 6.1). The queries were chosen in increasing order of complexity, to fully cover SemSur's capabilities.

Expert assessment: Hlomani and Stacey [12] proposed a metric suite for ontology evaluation, including accuracy, adaptability, clarity, cohesion, completeness, conciseness, consistency and coverage. Each metric is defined in the questionnaire to provide a clear description of it and to avoid ambiguity or misunderstanding. A total of 10 ontology engineering experts were recruited for this questionnaire[11]. Each expert had to give a percentage for each one of these metrics for the ontology.

6.1. Evaluation Results

All participants used digital libraries such as ACM DL or Web of Science and also use web search engines that index the full text or metadata of scholarly literature such as Google Scholar and sometimes using advanced search options and filters. The results retrieved this way were either out of scope for the query or merely related to the keywords, i.e., the search lacked semantics. All subjects unanimously agreed that the current way would not help them unless they explored the respective field more deeply, e.g., by reading survey articles on the topic. From analyzing the responses of the questionnaire participants, we made the following observations:

- Two-thirds of the participants found it difficult to obtain precise information using traditional tools,
- Around 72% of them believe that it is hard to get answers of queries like "*what are the approaches proposed to solve research problem using the traditional tools?*" ,
- Eight participants pointed out that SemSur would be very helpful for new researchers while ten agreed that SemSur would be *very helpful* for experienced researchers,
- Almost all participants believe that SemSur can help in either the decision of selecting relevant articles for a survey or finding the state-of-the-art approaches to compare their research with.
- About 94% of the participants indicate that the proposed approach saves a lot of time and effort.

It is worth mentioning that 12 participants are *very happy* to reuse or query SemSur in the future, four are *happy* and only one is *neutral*. *Regarding expert assessment*, the highest average percentage by experts is given to the *Clarity* (83%) which reflects how effectively the ontology communicates the intended meaning of the defined terms. The *Consistency*, i.e., are there any conflicts between ontology elements? , has the second highest percentage of 81% which demonstrates that the ontology does not include or allow for contradictions. Both *Conciseness* and *Accuracy* amount to 78%, which means that the asserted knowledge in the ontology agrees with the expert's knowledge about the domain. The average percentage of Adaptability, Cohesion, Completeness and Coverage are 67%, 77%, 74% and 74% respectively. Overall, eight participants taking the satisfaction questionnaire are strongly satisfied with SemSur and eight are satisfied. As anticipated, the evaluation results emphasize the validity of our ontology and show its potentially significant impact on scholarly communication.

[10] https://goo.gl/4SX2Bb
[11] https://goo.gl/8ybDf4

Table 5: Evaluation queries. A variable like X is a placeholder for any suitable text.

Query	Text
Q1	Publications title, year, keywords, authors and abstract financed by organization X?
Q2	Survey papers addressing problem X and its related problems?
Q3	Evaluation metrics, information assets, results and benchmarks are used to evaluate knowledge graph refinement frameworks?
Q4	Single-Author publications of person X proposed approach Y and addressed research problem Z?
Q5	Implementations for an approach X along with the platform, addressed problems, running OS, programming language?
Q6	Coauthors of person X sharing publications about research topic T along with their other publications?
Q7	Experiments design, goal and hypothesis used to evaluate implementations addressing research problem X?
Q8	Scientific articles that tackle the problem of generating RDF data from existing large quantities of data residing in relational databases?
Q9	Platforms used to implement approach X and information assets used for evaluation?
Q10	Motivations, approaches, experiment goal and frameworks for *"SPARQL query Federation"* and possible challenges and positive aspects?

7. Conclusion and Future work

In this article we have presented SemSur, a Semantic Survey Ontology, and an approach for creating a comprehensive knowledge graph representing research findings. SemSur breaks new ground in the studies of the transition from document-based to knowledge-based scholarly communication. We have defined an accompanying rule set for discovering new relationships and inferring new knowledge which does not explicitly exist in the knowledge graph. The results of our evaluation show that researchers agree that the traditional way of gathering an overview on a particular research topic is cumbersome and time-consuming. We have created instances of research findings of four survey articles on different fields of research to be able to determine if SemSur can answer typical information needs and give precise information. As an evidence, interviewed domain experts mentioned that it might be necessary to read and understand 30 to 100 scientific articles to get a proper level of understanding or an overview of a topic or sub-topics. Therefore, SemSur breaks new ground. As anticipated, SemSur enables successful retrieval of precise and accurate information about particular research aspects which potentially saves a lot of time and effort compared to traditional ways. In conclusion, SemSur can have a significant influence on the scientific community for both new and experienced researchers who want to write a survey or a literature review on a particular research topic.

To further our research, we are planning to integrate our methodology with the procedure of publishing survey articles can help to create a paradigm shift. As suggested by evaluation participants, we plan to further extend the ontology to cover fields of science like chemistry, physics and agriculture. For a more robust implementation of the proposed approach, we are planning to implement a user-friendly SPARQL auto-generation services for accessing metadata analysis for non-expert users and a web-based form to let researchers provide their findings. In addition, we are planning to implement an algorithm, inspired by the one proposed in [7], to identify relationships between individual research elements, e.g., benchmarks, and between authors, e.g., co-authorship. A more comprehensive evaluation of the services will be done after the implementation of the curation, exploration and discovery services. We will continue maintenance along three dimensions: 1) *Extension*: expand the inference rule set to cover further aspects and integrate more ontologies, thus enabling the ontology to capture research works in other fields of science, such as humanities or social sciences, 2) *Evaluation*: create more instances and more complex queries, and 3) *Improvement*: refine the inference rules and improve the coverage.

References

[1] Grigoris Antoniou and Frank Van Harmelen. *A semantic web primer*. MIT press, 2004.

[2] Alkyoni Baglatzi, Tomi Kauppinen and Carsten Keßler. *Linked science core vocabulary specification*. Tech. rep. available at, http://linkedscience.org/lsc/ns/, 2011.

[3] Paolo Ciccarese and Tudor Groza. 'Ontology of Rhetorical Blocks (ORB). Editor's Draft, 5 June 2011'. In: *World Wide Web Consortium. http://www. w3. org/2001/sw/hcls/notes/orb/(last visited March 12, 2012)* (2011).

[4] Alexandru Constantin et al. 'The document components ontology (DoCO)'. In: *Semantic Web* 7.2 (2016), pp. 167–181.

[5] Maciej Dabrowski, Macin Synak and Sebastian Ryszard Kruk. 'Bibliographic ontology'. In: *Semantic digital libraries*. Springer, 2009, pp. 103–122.

[6] Said Fathalla. 'Detecting Human Diseases Relatedness: A Spreading Activation Approach Over Ontologies'. In: *International Journal on Semantic Web and Information Systems (IJSWIS)* 14.3 (2018), pp. 120–133.

[7] Said Fathalla and Yaman Kannot. 'A Bidirectional-Based Spreading Activation Method for Human Diseases Relatedness Detection Using Disease Ontology'. In: *Conference on Computational Collective Intelligence Technologies and Applications*. Springer. 2017, pp. 14–23.

[8] Said Fathalla et al. 'Towards a Knowledge Graph Representing Research Findings by Semantifying Survey Articles'. In: *International Conference on Theory and Practice of Digital Libraries*. Springer. 2017, pp. 315–327.

[9] Mariano Fernández-López. 'Overview of methodologies for building ontologies'. In: (1999).

[10] Ernest Friedman-Hill. *JESS in Action*. Vol. 46. Manning Greenwich, 2003.

[11] Patrick Hayes and Brian McBride. 'RDF semantics'. In: *W3C recommendation* 10 (2004).

[12] Hlomani Hlomani and Deborah Stacey. 'Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey'. In: *Semantic Web Journal* (2014), pp. 1–5.

[13] Matthew Horridge et al. 'A practical guide to building owl ontologies using protégé 4 and co-ode tools edition1. 2'. In: *The university of Manchester* 107 (2009).

[14] Ian Horrocks et al. 'SWRL: A semantic web rule language combining OWL and RuleML'. In: *W3C Member submission* 21 (2004), p. 79.

[15] Senator Jeong and Hong-Gee Kim. 'SEDE: An ontology for scholarly event description'. In: *Journal of Information Science* 36.2 (2010), pp. 209–227.

[16] Markus Nentwig et al. 'A survey of current Link Discovery frameworks'. In: *Semantic Web* 8.3 (2017), pp. 419–436.

[17] Martin O'Connor and Amar Das. 'SQWRL: a query language for OWL'. In: *Proceedings of the 6th International Conference on OWL: Experiences and Directions-Volume 529*. CEUR-WS. org. 2009, pp. 208–215.

[18] Heiko Paulheim. 'Knowledge graph refinement: A survey of approaches and evaluation methods'. In: *Semantic web* 8.3 (2017), pp. 489–508.

[19] Silvio Peroni and David Shotton. 'FaBiO and CiTO: ontologies for describing bibliographic resources and citations'. In: *Web Semantics: Science, Services and Agents on the World Wide Web* 17 (2012), pp. 33–43.

[20] H Sofia Pinto, Asunción Gómez-Pérez and João P Martins. 'Some issues on ontology integration'. In: IJCAI and the Scandinavian AI Societies. CEUR Workshop Proceedings. 1999.

[21] H Sofia Pinto and JP Martins. 'Reusing ontologies'. In: *AAAI 2000 Spring Symposium on Bringing Knowledge to Business Processes*. Vol. 2. 000. Karlsruhe, Germany: AAAI. 2000, p. 7.

[22] Helena Sofia Pinto and João P Martins. 'Ontologies: How can they be built?' In: *Knowledge and Information Systems* 6.4 (2004), pp. 441–464.

[23] Mark Proctor. 'Drools: a rule engine for complex event processing'. In: *Proceedings of the 4th international conference on Applications of Graph Transformations with Industrial Relevance*. Springer-Verlag. 2011, pp. 2–2.

[24] Nur Aini Rakhmawati et al. 'Querying over Federated SPARQL Endpoints—A State of the Art Survey'. In: *arXiv preprint arXiv:1306.1723* (2013).

[25] Larisa N Soldatova and Ross D King. 'An ontology of scientific experiments'. In: *Journal of the Royal Society Interface* 3.11 (2006), pp. 795–803.

[26] Dimitrios-Emmanuel Spanos, Periklis Stavrou and Nikolas Mitrou. 'Bringing relational databases into the Semantic Web: A survey'. In: *Semantic Web* 3.2 (2012), pp. 169–209.

[27] York Sure et al. 'The SWRC ontology-semantic web for research communities'. In: *EPIA*. Springer. 2005, pp. 218–231.