Künstl Intell (2016) 30:117–130
DOI 10.1007/s13218-016-0424-1

TECHNICAL CONTRIBUTION

15 Years of Semantic Web: An Incomplete Survey

Birte Glimm[1] · **Heiner Stuckenschmidt**[2]

Received: 7 November 2015 / Accepted: 5 January 2016 / Published online: 23 January 2016

Abstract It has been 15 years since the first publications proposed the use of ontologies as a basis for defining information semantics on the Web starting what today is known as the Semantic Web Research Community. This work undoubtedly had a significant influence on AI as a field and in particular the knowledge representation and Reasoning Community that quickly identified new challenges and opportunities in using Description Logics in a practical setting. In this survey article, we will try to give an overview of the developments the field has gone through in these 15 years. We will look at three different aspects: the evolution of Semantic Web Language Standards, the evolution of central topics in the Semantic Web Community and the evolution of the research methodology.

Keywords Semantic Web · Ontologies · Knowledge representation · Survey

1 The Idea of the Semantic Web

In 1989, Tim Berners-Lee developed the vision for a World Wide Web in a document called "Information Management: A Proposal". Subsequently, he developed the three fundamental technologies that remain the foundation of today's Web: Uniform Resource Identifiers (URIs) for assigning unique identifiers to resources, the HyperText Markup Language (HTML) for specifying the formatting of Web pages, and the Hypertext Transfer Protocol (HTTP) that allows for the retrieval of linked resources from across the Web. Typically, the markup of standard Web pages describes only the formatting and, hence, Web pages and the navigation between them using hyperlinks is targeted towards human users (cf. Fig. 1).

In 2001, Tim Berners-Lee, James Hendler, and Ora Lassila describe their vision for a Semantic Web [5]:

> The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.GG

The core idea to realizing the Semantic Web is to no longer leave the semantics of (hyper-)links implicitly given, but to assign names/types also to the link between a source and a target resource (cf. Fig. 2). Each triple (source–link–target) can then be seen as an assertion that is presumed to be true (but may be a false statement). This idea of specifying basic statements in the form of triples led to the development of the Resource Description Framework (RDF), which is a data model for expressing descriptions of resources in the form of `subject predicate object` (short `s p o`) triples. A set of such triples is interpreted as a graph where `s` and `o` are nodes and `p` specifies the label for an edge from `s` to `o`. Starting from the source–link–target idea, in RDF source and target can also be arbitrary resources and not just Web pages and predicates are used to type the links.

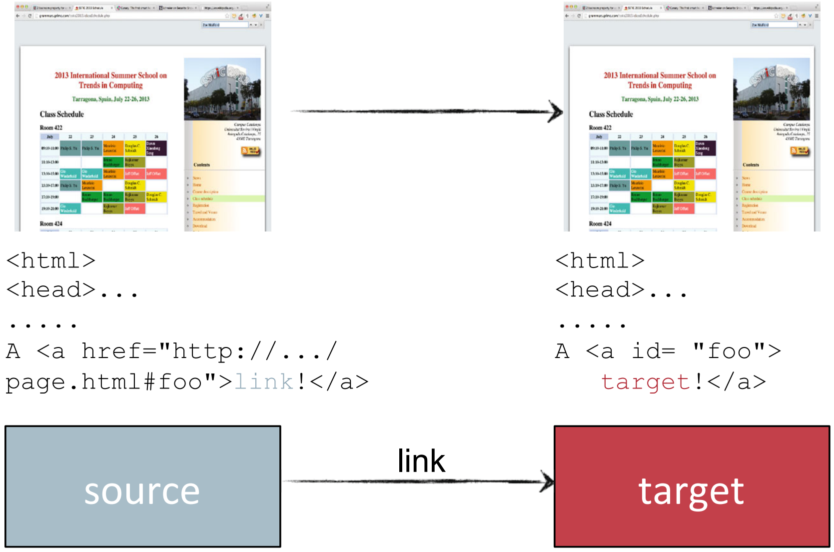
This basic idea of a data model for describing resources and relations between them lead to the development of several open standards to describe information over the last

✉ Birte Glimm
birte.glimm@uni-ulm.de

Heiner Stuckenschmidt
heiner@informatik.uni-mannheim.de

[1] Institute of Artificial Intelligence, University of Ulm, 89081 Ulm, Germany

[2] Chair of Artificial Intelligence, University of Mannheim, 68161 Mannheim, Germany



```
<html>
<head>...
.....
A <a href="http://.../
page.html#foo">link!</a>
```

```
<html>
<head>...
.....
A <a id= "foo">
   target!</a>
```

Fig. 1 Plain links between HTML pages

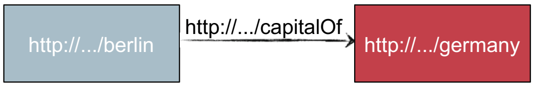


Fig. 2 Typed links between resources

15 years. The standards are clearly defined, flexible, extensible, and allow for deriving knowledge from the given information. In this survey article, we next summarize the evolution of the core Semantic Web standards. Section 3 then describes the evolution of central topics in the Semantic Web community and Sect. 4 gives insights into the evolution of the research methodology. Finally, we conclude this survey in Sect. 5.

2 Semantic Web Standards

Standards are an important basis for the Semantic Web to achieve interoperability across different systems and tools. The important standards for the Semantic Web are developed by the World Wide Web Consortium (W3C) whose founder and director Tim Berners-Lee originally envisioned an architecture that is based on a layered stack of technologies (cf. Fig. 3) informally called the "layer cake". The standards for these technologies have undergone a significant development over the last decade that we try to summarize in this section.

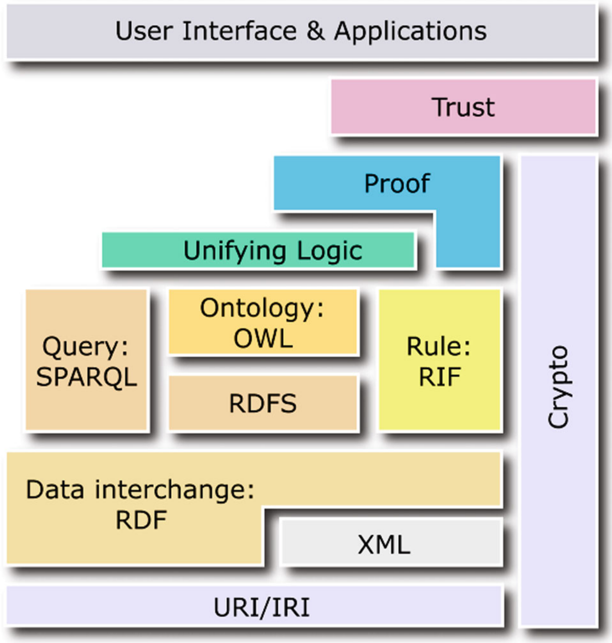


Fig. 3 The Semantic Web architecture Copyright 2001 World Wide Web Consortium (MIT, ERCIM, Keio, Beihang) http://www.w3.org/Consortium/Legal/2015/doc-license

2.1 The Resource Description Framework RDF

The most basic "layer" that is specific for the Semantic Web is the Resource Description Framework, for which the standardization process started in 1998 with the W3C recommendation ready in 2004 [27]. As outlined in the introduction, RDF is a data model to express descriptions of resources in the form of subject predicate object (s p o) triples, where a set of such triples is interpreted as a (labeled) graph. Subjects are resources given in the form of URIs or blank nodes, which only serve as an object identifier within the document without being globally valid. Predicates (also called properties) specify the relationship between the subject and the object and are given in the form of a URI. Objects are either URIs, blank nodes or (datatyped) literals. Datatypes are used by RDF in the representation of values such as integers, floating point numbers, or dates and most datatypes are taken from the XML Schema Datatypes specification [7]. The originally envisioned use of RDF was to specify meta data for Web resources, but this narrow scope was later widened to the encoding of structured, machine processable information in general.

RDF triples can be serialized in several formats of which the most widely used ones are the normative RDF/XML format [4], Turtle [2], and JSON[1] [16]. Figure 4 shows an example in Turtle syntax, where the first lines introduce (short) prefixes that are used to expand abbreviated URIs (called CURIEs [6]).[2] We use these prefixes also in the remainder of this article. In addition to prefixes, Turtle allows for several abbreviations for writing triples, e.g., instead of terminating a triple with a fullstop, one can use a semicolon (comma) to indicate that the next triple shares the same subject (subject and object), which is also used in the example. The first object in the example is a typed literal, which indicates that the string (i.e., the lexical value) “3517424” is to be interpreted as the integer value 3,517,424 as defined in the XML Schema Datatypes specification.

RDF provides a range of keywords some of which have a predefined semantics. For example, the keyword `rdf:type` used in the example indicates that the DBpedia resource for the city Berlin (`dbp:Berlin`) is of type (i.e., is an instance of the class) `dbo:Region` and `yago:CapitalsInEurope`. Hence, we have an implicit distinction between concrete elements (e.g., `dbp:Berlin`) and classes of elements (e.g., `yago:CapitalsInEurope`). While RDF allows for writing down facts with `rdf:type` as the only modeling construct, RDF Schema (RDFS) [11] extends the range of keywords with special semantics and allows for some forms of schema modelling. For example, one can state subclass or subproperty relationships and domains and ranges for properties, i.e., types that can be derived for the subject (domain) and object (range) of a triple related with the property. In the above example, the last triples makes use of the keyword `rdfs:subClassOf` and a reasoner that supports RDFS entailment would derive the triple `dbp:Berlin rdf:type dbo:PopulatedPlace`. Hence, with RDFS we can define important terms (for our application) and their relations to each other in the form of a so-called *ontology* or *terminology*. As such, RDFS is not only useful in the Semantic Web and is also used in other contexts where we can benefit from the use of standardized, formal languages that allow for inferring implicit knowledge by means of automated reasoning.

```
@prefix dbp: <http://dbpedia.org/resource/> .
@prefix dbo: <http://dbpedia.org/ontology/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf:
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .

dbp:Berlin dbo:populationTotal "3517424"^^xsd:integer ;
  owl:sameAs <http://sws.geonames.org/2950159/> ;
  rdf:type dbo:Region, yago:CapitalsInEurope .
dbo:Region rdfs:subClassOf dbo:PopulatedPlace .
```

Fig. 4 Extract from DBpedia in Turtle syntax

In 2014, the W3C standardized RDF 1.1 and RDFS, the successor versions of the initial standards from 2004 and we next outline the main differences. For a full list of differences we refer interested readers to a W3C Working Group Note [48].

Instead of URIs, the updated standard now uses Internationalized Resource Identifiers (IRIs). All literals must now have a datatype and an explicit list of RDF-compatible XSD datatypes is now part of the RDF 1.1 Concepts and Abstract Syntax standard [15]. Literals without a specified datatype, so-called simple literals, may now be supported as syntactic sugar for data values from the `xsd:string` datatype. Literals with a language tag (e.g., in ``` ``Example text''@en ``` the appended @ followed by the language tag `en` indicates that the string is an English language text) now have the (implicit) datatype IRI `rdf:langString`.

The query language SPARQL (see also Sect. 2.3) already allowed for the introduction of “names”, i.e., IRIs for RDF graphs. RDF 1.1 now also supports the notion of named graphs and additionally allows for the use of blank nodes as graph names. A collection of an unnamed, so-called default graph, and possibly several named graphs is then called an RDF dataset [49]. In order to serialize an RDF dataset one of the newly standardized syntaxes TriG, JSON-LD, and N-Quads [8, 13, 42] can be used. Figure 5 gives an overview of the old and new RDF serialization formats.

Regarding semantics, the new version only brings one significant change. Previously, it was not possible to express inconsistencies under RDF entailment. Only under RDFS entailment, one could express an inconsistency, by defining `rdf:XMLLiteral` as range for some property and by then using this property with an object that is a literal with an ill-formed XML fragment. In this case, it is not possible to find an interpretation that satisfies both triples. In RDF 1.1, graphs that contain an invalid literal for one of the RDF-recognized datatypes (e.g., ``` ``a''^^xsd:integer ```) are immediately inconsistent even under RDF entailment.

---

[1] http://json.org/.

[2] One can look-up typical prefixes at http://prefix.cc.

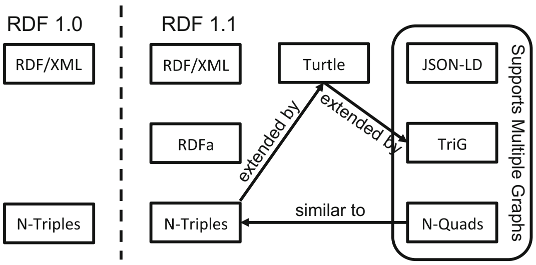


Fig. 5 RDF serialization formats

2.2 The Web Ontology Language OWL

As RDF, the first version of the Web ontology language OWL was standardized in 2004. The work on OWL is based on the results of the research projects DAML (DARPA Agent Markup Language) and Ontoknowledge, which defined the ontology languages DAML and OIL (Ontology-based Inference Layer). In 2009, OWL 2 followed with a second edition of the standard being released in 2012 with minor editorial changes compared to the 2009 release. OWL allows for a wide range of (schema) modeling constructs that have built-in semantics that can be implemented by automated reasoning procedures. Although OWL ontologies are RDF graphs [35], the RDF/XML and triple-oriented syntaxes are less convenient to state several complex modeling constructs and OWL 2 defines more convenient syntaxes to serialize OWL ontologies: the functional-style syntax [33], the OWL/XML syntax [31], and, as a Working Group Note and aimed at human-readability, the Manchester syntax [25].

Basic elements in OWL are *classes* (also called concepts), which represent sets of elements with common characteristics, *properties* (also called roles), which represent relations between pairs of elements, and *individuals*, which represent (named) elements in the domain. In contrast to RDF(S), OWL distinguishes between object and data properties, where the former relate two individuals/resources and the latter relate an individual with a data value. There are two semantics for OWL: one is an extension of the RDF(S) semantics, called OWL RDF-Based Semantics [39] and the other one is based on Description Logics, called OWL Direct Semantics [32]. For an introduction to Description Logics, we refer to the Description Logic Primer [28].

```
Prefix: : <http://www.loa-cnr.it/ontologies/
            DOLCE-Lite.owl#>
Class: :endurant
  Annotations:
    rdfs:comment "The main characteristic of endurants
                  is that all of them are independent
                  essential wholes. [...]"^^xsd:string
  SubClassOf:
    :participant-in some :perdurant,
    :spatio-temporal-particular,
    :specific-constant-constituent only :endurant,
    :part only :endurant
  DisjointWith:
    :quality, :perdurant, :abstract
```

Fig. 6 Axioms for the class endurant from the DOLCE upper level ontology in Manchester syntax

$$\begin{aligned} \text{endurant} \sqsubseteq\ & \exists\text{participant-in.perdurant} \sqcap \\ & \text{spatio-temporal-particular} \sqcap \\ & \forall\text{specific-constant-constituent.endurant} \sqcap \\ & \forall\text{part.endurant} \sqcap \\ & \neg\text{quality} \sqcap \neg\text{perdurant} \sqcap \neg\text{abstract} \end{aligned}$$

Fig. 7 Axioms for the class endurant from the DOLCE upper level ontology in DL syntax

Figures 6 and 7 show the axioms for the class *endurant* from the DOLCE[3] (Descriptive Ontology for Linguistic and Cognitive Engineering) upper level ontology in Manchester and DL syntax, respectively. Note that DL syntax does not support namespaces and the declaration of prefixes and the Manchester syntax has predefined prefixes for the common namespaces (`rdf`, `rdfs`, `owl`, `xsd`). This contributes to the fact that the RDF(S) and OWL keywords are not as visible, but the keywords `SubClassOf` and `DisjointWith` stand for `rdfs:subClassOf` and `owl:disjointWith`, respectively. The Manchester syntax makes use of OWL's annotations, which allow for "attaching" comments to axioms, where the comment itself uses the `rdfs:comment` keyword from the RDFS vocabulary. DLs do not support annotations. Further note that some of the stated superclasses make use of OWL's universal (`only`, ∀) and existential (`some`, ∃) quantifiers. For example, the (complex) class `:participant-in some :perdurant` (in DL syntax: ∃participant-in.perdurant) states that any instance of the class endurant is related to some (maybe not explicitly given) instance of the class perdurant via the (object) property participant-in. The expression `:specific-constant-constituent only :endurant` (in DL syntax: ∀specific-constant-constituent.endurant) requires that any instance of the class endurant is related only to instances of the class endurant via the (object) property specific-constant-constituent. Note that Fig. 6 does not display a proper OWL ontology since this would require the declaration of an ontology using the `Ontology:` keyword, possibly followed by an IRI for the ontology and a version IRI, which is used to identify the version of the ontology.

In order to illustrate how such axioms can be expressed as triples and which keywords from the OWL vocabulary

---

[3] http://www.loa.istc.cnr.it/old/DOLCE.html.

```
:endurant rdfs:subClassOf [
  rdf:type owl:Restriction ;
  owl:onProperty :participant-in ;
  owl:onClass :perdurant
] .
```

Fig. 8 Translation of the existentially quantified subclass axioms to Turtle

are used, Fig. 8 shows how the existentially quantified superclass of endurant is expressed in Turtle, where the opening square bracket introduces a blank node that is used as the subject of the following triples.

OWL reasoners consider the semantics of all special keywords (e.g., in subclass or disjointness statements) and check whether an ontology is consistent (i.e., free of logical contradictions), compute the subsumption hierarchy (i.e., explicitly stated and logically following subclass relationships), or derive other logically entailed facts. Hence, reasoners help in finding modeling errors in an ontology that manifest themselves in unwanted subsumptions or even an inconsistency. Furthermore, debugging is facilitated since reasoners can compute which axioms cause an (unwanted) entailment or an inconsistency.

Apart from the new syntaxes and a clear definition of the two semantics, OWL 2 adds several other features [22]. Ontologies can, for example use an import directive to include also axioms and facts from other ontologies. Regarding modeling constructs, some syntactic sugar (e.g., negative property assertions) as well as new constructors for properties (e.g., for declaring a property as symmetric or disjoint to another one or for defining property chain inclusions) and classes (e.g., qualified cardinality restrictions to require a minimal, maximal or exact number of instances of a given class to which an individual is related via a given property) have been added. Furthermore, the datatype capabilities have been significantly extended by allowing for custom datarange definitions based on existing datatypes using factes [7] and logical constructors (conjunction, disjunction, negation). The following two axioms in Manchester syntax illustrate these features by defining a custom datatype based on `xsd:integer` (abbreviated to `integer` in the Manchester syntax) that is then used together with another custom data range in defining the class `NonAdult`:

```
Datatype: :TeenAge
  EquivalentTo: integer [> 13, < 19]
Class: :NonAdult
  EquivalentTo: :hasAge some
    (:TeenAge or nonNegativeInteger [<= 13])
```

OWL 1 defined three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full. The worst-case complexity of reasoning in these sub-languages is ExpTime, NExpTime, and undecidable, respectively, which illustrates that OWL Lite is not really (computationally) light-weight. OWL 2 addresses this by defining three so-called profiles, which are tractable for certain reasoning tasks.

1. OWL EL is based on the Description Logic $\mathcal{EL}$++ [1], which captures, for example, many large bio-medical ontologies. The computation of the class hierarchy (all subclass relationships) can be implemented efficiently with polynomial worst-case complexity.
2. OWL QL allows for answering conjunctive queries in AC^0 (data complexity, i.e., with respect to the size of the assertions/facts), while being able to capture the main features necessary to express conceptual models such as UML class diagrams and ER diagrams. The profile is based on the Description Logic DL-Lite [12] and is designed so that data (assertions/facts) can be stored in a standard relational database system and queries can be answered via a simple rewriting mechanism, i.e., by rewriting the query into an SQL query that captures the semantics of the schema axioms without any changes to the data.
3. OWL RL is defined such that reasoning can be implemented by standard rules engines (as RDF(S) reasoning) in polynomial time. The design of OWL 2 RL was inspired by Description Logic Programs [23] and pD* [44].

OWL EL and QL are defined for the Direct Semantics, whereas OWL 2 RL works with both the Direct and the RDF-Based Semantics. The worst-case complexities of the different OWL sub-languages are illustrated in Fig. 9.

2.3 The Query Language SPARQL

SPARQL stands for SPARQL Protocol And RDF Query Language and is a W3C specification since 2008 with an extension to SPARQL 1.1 in 2013. The initial specification consists of three parts: the query language [38], the XML result format [3], and the query protocol for the transmission of query and result [14]. Version 1.1 extends the query language and introduces several new features, which we try to summarize in this section.

As a simple example, consider the following SPARQL query, which is to be evaluated over the DBpedia data from Fig. 4.

```
@PREFIX dbo: <http://dbpedia.org/ontology/> .
SELECT ?city ?population WHERE {
  ?city dbo:populationTotal ?population .
}
```

While SPARQL is similar to SQL, it is a native Semantic Web language and, as such, it allows, for example, the

---

[3] http://www.loa.istc.cnr.it/old/DOLCE.html

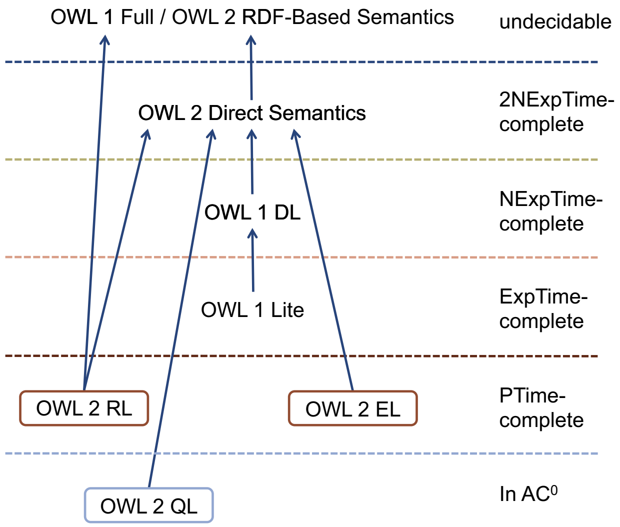


Fig. 9 Worst-case complexities of the different OWL sub-languages

declaration of prefixes. After the `SELECT` keyword, we list the variables, prefixed with `?` or `$`, that are to be selected (`*` selects all variables). People familiar with SQL might miss the `FROM` clause, which can be omitted in SPARQL since queries are assumed to be evaluated over an RDF data set (cf. Sect. 2.1) and queries without a from clause are evaluated over the default graph of the data set. The `FROM NAMED` keyword can be used to evaluate queries over a named graph of the data set. After the `WHERE` keyword, the query pattern is given, enclosed in curly brackets. The simplest query pattern, called a basic graph pattern (BGP), consists of triples where subject, predicate, and object can be replaced by variables. Answers to such a query are obtained by mapping the BGP onto the queried graph with variables acting as wild cards. The so obtained bindings might be further processed, e.g., by projecting out variables as required by the select clause.

More complex BGPs can be formed by using the keyword `UNION`, which allows for specifying alternative parts for a pattern, `OPTIONAL` to enrich results with optional mappings, and `FILTER` followed by a filter expression that evaluate to truth values (and possibly errors). Many filter functions are partly taken from the XQuery/XPath-standard for XML [29]. Assume, for example, the modification of the query above:

```
@PREFIX dbp: <http://dbpedia.org/resource/>
@PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT DISTINCT ?city ?population WHERE {
  { ?city dbo:populationTotal ?population . }
  UNION
  { ?city dbp:population ?population . }
  FILTER (?population > 10000)
} ORDER BY ?population
```

We also select cities that use `dbp:population` instead of `dbo:populationTotal`, where `DISTINCT` eliminates possible duplicates and the filter eliminates results where the population is smaller than 10,000. The `ORDER BY` keyword is a so-called solution modifier, which requires the results to be returned ordered in ascending (default) order by the population. Apart from `SELECT` queries, SPARQL also supports other kinds of queries, e.g., query results can be returned in the form of an RDF graph using the `CONSTRUCT` keyword followed by a pattern that is instantiated with the results. In order to formally describe the semantics of queries, the SPARQL Query specification describes the translation to and evaluation of algebra objects.

SPARQL 1.1 adds not only new features to the query language, but also several new specifications. We refer to the overview [45] for a detailed description and only summarize the changes here. Regarding the query language, important new features are

- expressions in the select or the where clause to compute values, e.g., `(?price * ?amount) AS ?sum),`
- aggregates, e.g., to count the number of results or to compute average values
- property paths that allow for regular expressions over properties, e.g., we can query for ancestors with the property path expression

  `?ancestor(ex:motherOf|ex:fatherOf)+ <#me>`

  by following a path of arbitrary length (+) over `ex:motherOf` or (|) `ex:fatherOf` labeled edges,
- negated patterns (`MINUS` or `FILTER NOT EXISTS`).

SPARQL 1.1 further adds a range of new features in new specifications: *SPARQL 1.1 Update* [20] extends the query language to allow for the manipulation of graphs or graph content. *SPARQL 1.1 Entailment Regimes* [21] redefine the evaluation of BGPs such that entailment relations are used instead of subgraph matching to define the results. Supported entailment relations are RDF, RDFS, D (RDFS with extended datatype support), OWL Direct and RDF-Based Semantics, and RDF extended with RIF rules (cf. Sect. 2.4). SPARQL endpoints that support entailment consider the special semantics of the supported vocabularies. For example, if the endpoint employs RDFS entailment and one queries for instances of a class, then also asserted instances of subclasses are in the result. *SPARQL 1.1 Service Descriptions* [47] provide a vocabulary and method for describing SPARQL endpoints such that clients/users can request information about the SPARQL service, e.g., supported extension functions, used data set or supported inference mechanisms. *SPARQL 1.1 Federated Query* [37] defines how queries distributed over different SPARQL endpoints can be executed, which allows for directly merging data within a query that is distributed across the Web. *The XML results* format has only slightly been adapted [24], but two new specifications define a serialization of results in JSON [41] and comma/tab separated value (CSV/TSV) format [40]. *The SPARQL 1.1 Protocol* [19] has been extended to also cover the SPARQL UPDATE operations. *The SPARQL 1.1 Graph Store HTTP Protocol* [34] describes the use of HTTP operations for the purpose of managing a collection of graphs in the REST architectural style.

2.4 The Rule Interchange Format

The Rule Interchange Format (RIF) [26] is a standard for exchanging rules among rule systems, in particular among Web rule engines. In contrast to the standards described above, RIF is not a rule language, but an exchange format due to the many popular paradigms for using rules that can hardly be captured by a single language. It was standardized in 2010 with a second edition that incorporates editorial changes in 2013. We just summarize RIF very briefly and refer interested readers to the RIF primer [30] for a more detailed introduction.

The most basic RIF language is RIF-Core [9], which is augmented by a set of datatypes and built-in functions and predicates (RIF-DTB) that can be used when writing rules [36]. All further RIF dialects are an extension of RIF-Core plus DTB. The Basic Logic Dialect (RIF BLD) [10] extends RIF-Core with function symbols and equality, which implies that rule engines can no longer guarantee termination. The RIF Production Rule dialect [18] allows for specifying production rules with an operational semantics instead of the model-theoretic semantics used by BLD. Extensions to RIF Core are the support of priorities, i.e., one can specify that some rules are considered before others, negation in the rule condition/if clause, and knowledge base modification, e.g., a rule application might add statements to or retract statements from a knowledge base.

The standard syntax for RIF is a verbose XML syntax. For better readability, Fig. 10 gives an example using the Mixed Presentation Syntax that combines features of the Abstract and the Presentation Syntax used as compact syntaxes for PRD and BLD, respectively. As in previously presented examples, one can declare prefixes in RIF to abbreviate IRIs. The actual rules are embedded within a document declaration and, as also OWL, RIF supports a mechanism to import rules from other documents. Groups are organisational structures for humans to keep rules together that share some commonalities. In the example, we have two groups of which the first contains a mapping rule and the second contains a rule that implements (part of) the RDFS semantics. As in SPARQL, variables are prefixed with a question mark and both rules apply universally (`Forall`). Note that within the if clause of a rule one can also use existentially quantified variables. Given the first rule, a (declarative) rule engine would derive, for any instance (`rdf:type`) of the (DBpedia) class `dbo:Region`, that the individual is also an instance of the class `ex:Area`, where `ex` is a prefix for some imaginary, example URL. Hence, with such a rule, we can map from one vocabulary to another. The second rule implements the RDFS semantics that instances of a class are also instances of the superclass of that class. While it is possible to implement RDFS or OWL RL entailment via a RIF rule set, the RIF RDF and OWL Compatibility standard [17] directly defines the semantics of RIF rule sets in combination with RDF(S) graphs or OWL ontologies.

```
Prefix(rdfs <http://www.w3.org/2000/01/rdf-schema#>)
Prefix(rdf
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>)
Prefix(dbo <http://dbpedia.org/ontology/>)
Prefix(ex <http://example.org/myOntology/>)

Document(
  Group(
    Forall ?individual (
      If   rdf:type(?individual dbo:Region)
      Then rdf:type(?individual ex:Area)
    )
  )
  Group(
    Forall ?individual ?sub ?super (
      If   And(rdf:type(?individual ?sub)
               rdfs:subClassOf(?sup ?super))
      Then rdf:type(?individual ?super)
    )
  )
)
```

Fig. 10 RIF example in Mixed Presentation Syntax

3 Semantic Web Research Topics

Like any other research field, the Semantic Web has gone through a number of changes in the focus of the research, as the field became more mature. In this section, we try to sketch some of these developments of the topics investigated using empirical information from the International Semantic Web Conference as a basis.

3.1 Important Subtopics

The original idea of the Semantic Web was pretty much focussed on the development of representation languages for ontologies and factual knowledge on the Web that should provide the basis for answering questions that need information from different Web pages. Over time the focus of Semantic Web research has been significantly extended to other topics. In addition to knowledge representation and reasoning topics from related communities, in particular from Databases, Data Mining, Information Retrieval and Computational Linguistics. In the following, we describe some major topics that have been the subject of research in the Semantic Web community.

Language Standards and Extensions The development of standardized knowledge representation languages was the starting point of Semantic Web research. Languages like DAML, OWL and RDF, but also representation languages for services such as DAML-S, OWL-S and WSMO were developed and various extensions were proposed, only some of which actually made it into the official language standard. Further, researchers discussed the use of other existing languages like XML as a basis for the Semantic Web.

Logic and Reasoning Most of the language standards proposed for the Semantic Web are based on some formal logic. Thus extending existing logics and reasoners to completely cover the respective standards as well as the development of scalable and efficient reasoning methods have been in the focus of research form the beginning. Over the years the scope has been extended to genuine research on logics for reasoning about data, including non-standard reasoning methods and combinations of logics with non-logical reasoning paradigms.

Ontologies and Modelling The existence of language standards is necessary for Semantic Web applications, but it does not enable people to build the right models. Therefore, there has been work on ontological modeling that covers topics like modeling patterns, best practices and lessons learned as well as reports about large scale modeling efforts for instance in the biomedical domain. This work is mostly inspired by classical knowledge acquisition that was one of the first drivers of Semantic Web research.

Semantic Web Services While ontologies typically capture static knowledge, the description and use of dynamic Web elements, i.e., Web Services, has also been addressed in the context of Semantic Web. Besides fundamental work on describing services, typical topics are Web Service search and matching as well as automatic composition and execution of the services as part of a dynamically generated application.

Linked Data At some point it became clear that a top-down approach to the vision of the Semantic Web that is based on the creation of expressive ontologies and detailed Semantic modeling has too high ramp up costs to be attractive to a wider audience. As a reaction, linked data has been proposed as a bottom-up approach, where data is converted into Semantic Web standards with minimal ontological commitment, published and linked to other data sources. Corresponding work typically deals with tools and guidelines for publishing and linking data as well as reports on data publication projects.

Matching and Integration The integration of data and knowledge from different sources has remained a central topic both in the more classical view of Semantic Web research, where ontology matching is a dominant topic, as well as in the linked data area where the focus is more on matching individual data entries to establish links between datasets. Both can be seen as extensions of work in the database community on schema matching and record linkage. The use of logical semantics and reasoning to support the matching process can be seen as a unique contribution of Semantic Web research in these areas.

Query Processing With the availability of large RDF datasets created as linked data, the problem of efficiently accessing these datasets has become more important over the years. In particular, query processing for the SPARQL query language has become a dominant topic in Semantic Web research. While the relevant aspects a roughly the same as for query processing in relational databases, the nature of linked data has led to a stronger focus on distributed data storage and federated query processing and on the use of expressive schema information in terms of ontologies.

Security, Trust and Provenance It has been assumed from the beginning that in an open information environment like the Web the reliability of information will become an issue. In this context mechanism have been investigated for describing the origin (provenance) and the reliability (trust) and the protection (security) of data. Although there is continuous work on these aspects, none of them have really become a major topic beyond some specific application areas.

Knowledge Extraction and Discovery Information extraction from text has played a role in knowledge acquisition for a long time. With the change of focus of Semantic Web research from abstract models to large datasets, the importance of Data Mining and Machine Learning has increased significantly. Especially the generation of open domain datasets and knowledge bases from the Web has become a central topic in the community, also fueled by the large scale investments of Google and Microsoft in the creation of general purpose knowledge bases from Web data.

Search, Retrieval and Ranking The development of better search algorithms for finding the right information, which is clearly an important goal of Google's activities on knowledge extraction and integration, has—strangely enough—so far not been in the center of attention in the Semantic Web community. There has been some work on specific aspects like domain specific search engines, natural language access to linked data sets and so on; Web search as such, however, is something that has not been taken up as a task by the community so far.

User Interfaces and Annotation Despite all automation efforts, the human in the loop has turned out to not only be indispensable in some use cases, it has also turned out that making use of human input can be more effective. This observation has led to work on user interfaces for Semantic Web data and as a means for getting users involved in data annotation tasks.

Applications While being independent of a particular application domain, over the years a number of domains have being identified where Semantic Web technologies provide a direct benefit. Examples are the medical domain, research in bio-informatics as well as libraries and information science. In these and other less prominent domains, applications of Semantic Web technologies are also an important area of work.

3.2 Importance over Time

These topics have been more or less important in the community during the past 15 years. We have analyzed the papers accepted as full papers in the main research track of the International Semantic Web Conference (ISWC) from 2002 to 2014 and assigned them to one of the topics listed above. Over the whole period, most papers were concerned with query processing (91) followed by Logic and Reasoning (85) and 'Knowledge Extraction and Discovery' (75). With 64 papers 'Matching and Integration' is also well represented. These four topics are by far the most important ones. They are followed by 'Applications/Other' and 'Ontologies and Modeling' with more than 40 papers each. 'Linked Data', 'Web Services' and 'User Interfaces and Annotation' appeared between 30 and 35 times. The remaining topics 'Search, Retrieval and Ranking', 'Security, Trust and Provenance' and 'Language Standards and Extensions' each appeared about 25 times and are the less dominant topics.

Figure 11 shows the fraction of papers devoted to each topic for all ISWC Conference since 2002. It reveals some interesting developments in the focus of the community over the years. First of all, we can observe that some topics that have been at the core of the community initially gradually lost importance. In particular, this observation applies to the topics 'Language Standards and Extensions', 'Web Services' and 'Ontologies and Modeling'. The development and extension of standards were mostly moved to the respective standardization bodies at the W3C and were no longer discussed in scientific papers, which are now more focussed on the underlying formalisms and algorithms. Web Services are still an active area of research, which, however, is no longer strongly associated with rich semantic annotations and a high degree of automation as it was envisioned in the early days of Semantic Web research. In the case of 'Ontologies and Modeling', it seems that there is not that much emphasis on general principles any more as have been discussed in the AI community since the Eighties. Much of the work on building ontologies is now done and also published in the respective application domains, in particular in the medical domain that still has a very active medical ontologies community. Some of the topics show a relatively constant occurrence in the ISWC conferences. Amongst these topics there are some less prominent ones like 'Search, Retrieval and Ranking', 'Security, Trust and Provenance' and 'User Interfaces and Annotation' that have a constant but rather small share of the publications. However, there are also some more prominent topics whose share has stayed more or less constant over the years. In particular the topic 'Matching and Integration' shows such behavior. Although the fraction of papers on 'Logic and Reasoning' has shrunk in the last two years, overall it shows a rather stable occurrence in the conference. Two topics stick out that have become more important over the years. These are 'Query Processing' and 'Knowledge Extraction and Discovery'. This shows the increasing involvement of the Databases and Data Mining community in Semantic Web research.

4 Development of the Research Methodology

In a paper published at ISWC 2013 [43], we have investigated the research methodology of the Semantic Web Community by analyzing the nature of ISWC publications as well as the experimental work done. For this purpose, we annotated the ISWC publications with respect to the type of research using a classification proposed by Tichy and others [46] for classifying computer science research:

1. Formal Theory Papers whose main contributions are formal propositions, e.g., lemmata and theorems and their proofs.
2. Design and Modeling Papers whose main contributions are systems, techniques (e.g., algorithms) or models whose claimed properties cannot formally be proven.
3. Empirical Work/Hypothesis Testing Papers that collect, analyze and interpret observations about known designs, systems, models, or hypotheses.
4. Other Papers that do not fit the other categories (e.g., surveys).

As expected, the great majority of work on the Semantic Web falls into the category 'Design and Modelling' (80.8 %), followed by 'Formal Theory' (11.2 %) and only 5.4 % empirical work.

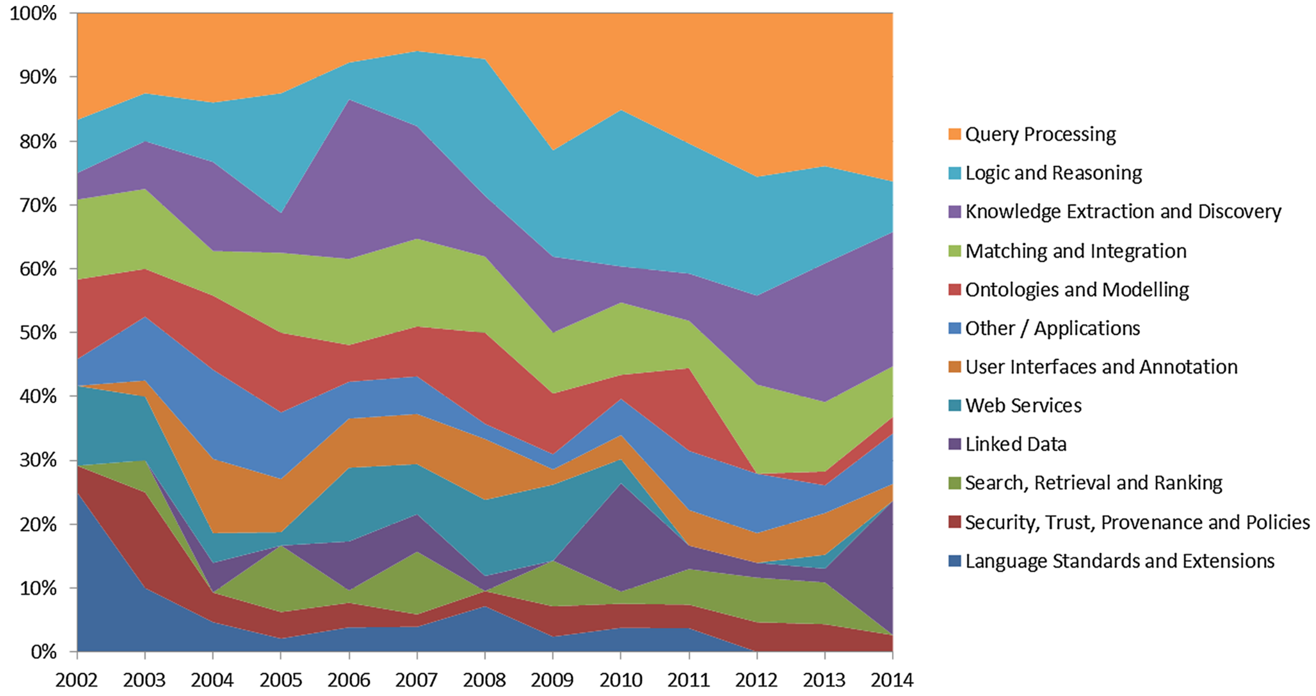


Fig. 11 Fraction of papers at the International Semantic Web Conference by topic

Focussing on the 'Design and Modelling' type papers, we investigated the importance and quality of experimental work used to support the claims about the system or model proposed in the paper. As proposed by Tichy et al. [46], we used the number of pages devoted to the description of the experiments as an indicator of the importance of the experimental work. In particular, we analyze how the papers distribute across the subcategories defined by the fraction of the pages devoted to the description of experimental work (0 %, (0–10 %], (20–50 %], >50 %) and look at the development of experimental work over time by plotting the distribution of papers across all categories over the past eleven years. We also look at the average number of pages devoted to experimental work in the different years and compute the correlation between year of publication and number of pages.

Figure 12 shows that an increase in the importance can be observed. It shows a standard box-plot for the relative number of experiment pages for Category 2 (Design and Modeling) papers. We identified a trend of growing importance of experiments over time. With the exception of 2010, the median is constantly rising up to 25 % in 2012. Measuring this trend in figures, the Spearman Correlation Coefficient is statistically significant ($r_S(402) = .49,\ p < .000$). This means that the importance of experimental work was rather low in the early years of the ISWC conference. This is not uncommon for new fields of research, as first, the principled ideas have to be laid out and basic ideas have to be tested in prototypical form. Only later, when the field is more established and the problems are better understood, systematic experiments become the standard way of validation.

Further, we annotate all papers of Category 2 with the following information about the nature of the experiments.

Standard used for Comparison Does the paper report about *different settings* or the system or method? Are results compared against *existing baselines*? Are results compared against the results of *other systems*? The latter includes both indirect comparisons against results reported in other papers and direct comparisons obtained by executing the other system as part of the experiments.

Datasets used Has *one dataset* been used or have *several datasets* been used within the experiments? Has the dataset been *self-created* by the authors for the purpose of conducting the experiments or is it *externally provided*?

We use this information as an indication of the quality of the experimental design, assuming that an ideal experimental design will compare a proposed system against other leading systems or at least sensible baselines using several datasets with different characteristics. One can argue about whether externally provided datasets should be preferred over self-created ones, in many cases externally provided datasets are publicly accessible benchmarks that support the comparison with other systems, which we consider desirable.

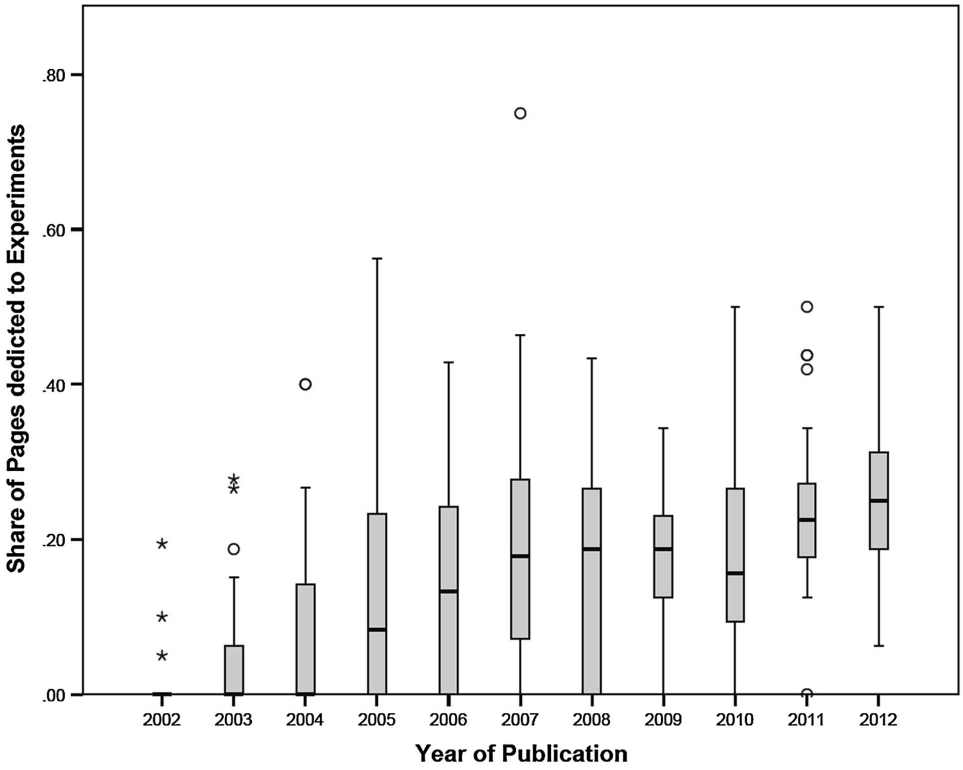


Fig. 12 *Box-plot* showing the relative number of pages of Category 2 (Design and Modeling) papers by year of publication. The median starting at 0 % in 2002 increases constantly (with the exception of 2010) over time, reaching its top of 24 % in 2012. The second/third quartile, denoted by the *box*, varies, but is since 2009 clearly above zero. Outliers are displayed as *circles/stars* (taken from [43])

The variables Several and Other can be interpreted as indicators for the universal validity of the reported results. The variables BaseDiff and Sys indicate whether the authors informed the reader on the performance (e.g., runtimes), quality (e.g., precision), or usability compared to alternative approaches. Without such a comparison, it is hardly possible to draw any conclusions related to the improvements made.

The results of our analysis are shown in Fig. 13, where we depicted the counts for all four variables with respect to Category 2 papers. Figure 13 reveals a clear trend. The quality of experimental work is increasing over time with respect to each variable. In 2003 only a minor share of all papers had a positive characteristic in one of the four variables, while in 2012 more than 50 % of all papers had a positive characteristic in three of four variables. However, only 33 % of all papers in 2012 compared their results against other systems (Sys). While this is an improvement compared to the previous years, there are still many papers that do not compare their results against other systems. We computed also the correlation between the year of publication and the four quality measures using Spearman's rank correlation coefficient. We find that all variables show positive and statistically significant correlations with the year of publication (r_S between .36 and .46, $p \leq .000$).

Our observations can be explained by two factors. One factor might be an increasing awareness of the importance attributed to experimental work. Another factor might be the general development of the community. What has been a novel area of research ten years ago, might have become an established research area associated with well-defined problems, commonly accepted formats, well-known datasets and accepted benchmarks. Obviously, both factors go hand in hand, resulting in the positive trend that we reported in the evaluation.

5 Conclusions

Over the past 15 years, the Semantic Web has established itself as a research area in its own rights. Based on the development of agreed standards for representing and accessing data on the web, the filed has generated a unique set of methods that address the specific needs of semantic information processing in an open and distributed environment. While most methods that are currently investigated have their origin in other more established research fields, in particular knowledge representation, database systems and knowledge discovery, the field meanwhile has made unique contributions to the state of the art and also starts to have impact on the research topics in these areas. The Semantic Web as a research field is still characterized by a mix of technologies and methods from different fields and by the application rather than by a unique set of

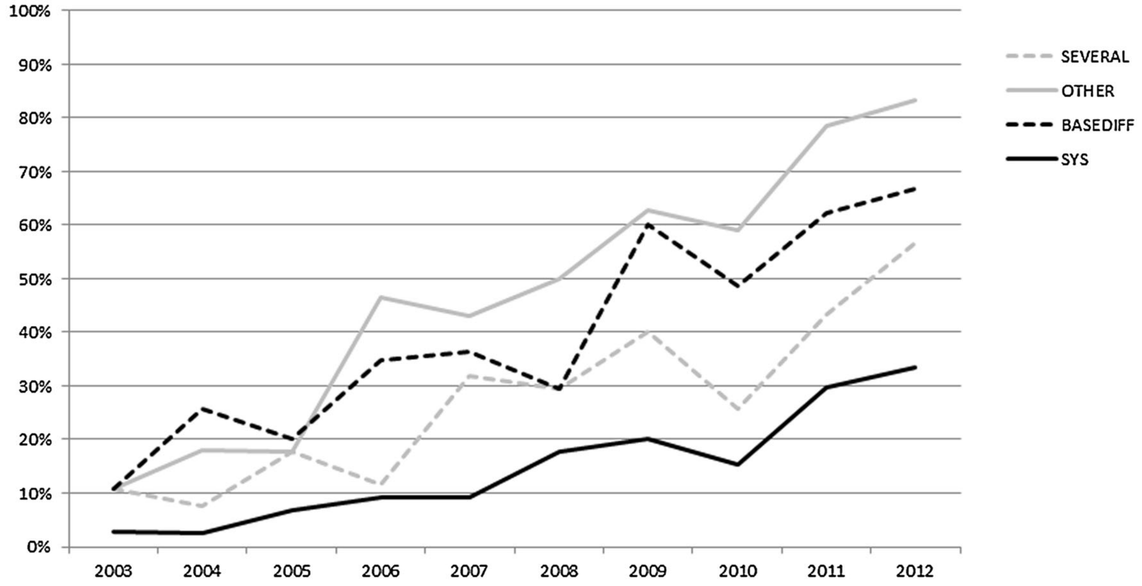


Fig. 13 Development of relative share of Category 2 (Design and Modeling) papers complying to different evaluation quality indicators over time. While all indicators start at a low level of ≤11 % in 2003 and rise with the years, we found the usage of externally provided datasets (Other) to increase the most. Nevertheless, even in 2012 only about on third of all papers compare themselves to other existing systems (Sys)

theories and tools. As we have seen in the last section, the growing maturity of the filed can be observed through the improvement of solid, experiment-based research methodologies that have been established over the years.

In summary, we can conclude that the Semantic Web has developed from an innovative idea to an established area of research unlike other topics that started with similar ambitions but meanwhile have disappeared again.

References

1. Baader F, Brandt S, Lutz C (2008) Pushing the EL envelope further. In: Clark K, Patel-Schneider PF (eds) In: Proceedings of the OWLED 2008 DC workshop on OWL: experiences and directions
2. Beckett D, Berners-Lee T. Turtle—Terse RDF Triple Language. In: W3C Team submission. W3C, 14 Jan 2008. http://www.w3.org/TeamSubmission/turtle/
3. Beckett D, Broekstra J. SPARQL query results XML format. In: W3C recommendation. W3C, 15 Jan 2008. http://www.w3.org/TR/rdf-sparql-XMLres/
4. Beckett D, McBride B. RDF/XML syntax specification. In: W3C recommendation. W3C, 10 Feb 2004. http://www.w3.org/TR/rdf-syntax-grammar/
5. Berners-Lee T, Hendler J, Lassila O (2001) The semantic web. Sci Am 284(5):34–43
6. Birbeck M, McCarron S: CURIE Syntax 1.0—a syntax for expressing compact URIs. In: W3C Working Group Note. W3C, 16 Dec 2010. http://www.w3.org/TR/curie
7. Biron PV, Malhotra A. XML schema part 2: datatypes second edition. In: W3C recommendation. W3C, 28 Oct 2004. http://www.w3.org/TR/2004/REC-xmlschema-2-20041028/
8. Bizer C, Cyganiak R. RDF 1.1 TriG—RDF dataset language. In: Carothers G, Seaborne A (eds) W3C recommendation. W3C, 25 Feb 2014. http://www.w3.org/TR/2014/REC-trig-20140225/
9. Boley H, Hallmark G, Kifer M, Paschke A, Polleres A, Reynolds D. RIF core dialect (second edition). In: W3C recommendation. W3C, 5 Feb 2013. http://www.w3.org/TR/2013/REC-rif-core-20130205/
10. Boley H, Kifer M. RIF basic logic dialect (second edition). In: W3C recommendation. W3C, 5 Feb 2013. http://www.w3.org/TR/2013/REC-rif-bld-20130205/
11. Brickley D, Guha RV, McBride B. RDF vocabulary description language 1.0: RDF schema. In: W3C recommendation. W3C, 10 Feb 2004. http://www.w3.org/TR/2004/REC-rdf-schema-20040210/
12. Calvanese D, De Giacomo G, Lembo D, Lenzerini M, Rosati R (2007) Tractable reasoning and efficient query answering in description logics: the DL-lite family. J Autom Reason 39(3):385–429
13. Carothers G. RDF 1.1 N-quads—a line-based syntax for RDF datasets. In: W3C recommendation. W3C, 25 Feb 2014. http://www.w3.org/TR/2014/REC-n-quads-20140225/
14. Clark KG, Feigenbaum L, Torres E. SPARQL protocol for RDF. In: W3C recommendation. W3C, 15 Jan 2008. http://www.w3.org/TR/rdf-sparql-protocol/
15. Cyganiak R, Wood D, Lanthaler M. RDF 1.1 concepts and abstract syntax. In: W3C recommendation. W3C, 25 Feb 2014. http://www.w3.org/TR/rdf11-concepts/
16. Davis I, Steiner T, Le Hors AJ. RDF 1.1 JSON alternate serialization (RDF/JSON). In: W3C Working Group Note. W3C, 07 Nov 2013. http://www.w3.org/TR/rdf-json/
17. de Bruijn J. RIF RDF and OWL compatibility (second edition). In: W3C recommendation. W3C, 5 Feb 2013. http://www.w3.org/TR/2013/REC-rif-rdf-owl-20130205/
18. de Sainte Marie C, Hallmark G, Paschke A. RIF production rule dialect (second edition). In: W3C recommendation. W3C, 5 Feb 2013. http://www.w3.org/TR/2013/REC-rif-prd-20130205/

19. Feigenbaum L, Williams GT, Clark KG, Torres E. SPARQL 1.1 protocol. In: W3C recommendation. W3C, 21 Mar 2013. http://www.w3.org/TR/2013/REC-sparql11-protocol-20130321/
20. Gearon P, Passant A, Polleres A. SPARQL 1.1 update. In: W3C recommendation. W3C, 21 Mar 2013. http://www.w3.org/TR/2013/REC-sparql11-update-20130321/
21. Glimm B, Ogbuji C. SPARQL 1.1 entailment regimes. In: W3C recommendation. W3C, 21 Mar 2013. http://www.w3.org/TR/2013/REC-sparql11-entailment-20130321/
22. Golbreich C, Wallace EK. OWL 2 web ontology language—new features and rationale (second edition). In: W3C recommendation. W3C, 11 Dec 2012. http://www.w3.org/TR/2012/REC-owl2-new-features-20121211/
23. Grosof BN, Horrocks I, Volz R, Decker S. Description logic programs: combining logic programs with description logic. In: Proceedings of the 12th international conference on world wide web (WWW 2003). ACM Press and Addison Wesley, pp 48–57 (2003)
24. Hawke S, Beckett D, Broekstra J. SPARQL query results XML format (second edition). In: W3C recommendation. W3C, 21 Mar 2013. http://www.w3.org/TR/2013/REC-rdf-sparql-XMLres-20130321/
25. Horridge M, Patel-Schneider PF. OWL 2 web ontology language—Manchester syntax. In: W3C Working Group Note. W3C, 27 Oct 2009. http://www.w3.org/TR/2009/NOTE-owl2-manchester-syntax-20091027/
26. Kifer M, Boley H. RIF overview (second edition). In: W3C Working Group Note. W3C, 5 Feb 2013. http://www.w3.org/TR/2013/NOTE-rif-overview-20130205/
27. Klyne G, Carroll JJ. Resource description framework (RDF): concepts and abstract syntax. In: W3C recommendation. W3C, 10 Feb 2004. http://www.w3.org/TR/rdf-concepts/
28. Krötzsch M, Simancik F, Horrocks I. A description logic primer. CoRR arXiv:1201.4089 (2012)
29. Malhotra A, Melton J, Walsh N. XQuery 1.0 and XPath 2.0 functions and operators. In: W3C recommendation. W3C, 23 Jan 2007. http://www.w3.org/TR/2007/REC-xpath-functions-20070123/
30. Morgenstern L, Welty C, Boley H, Hallmark G. RIF primer (second edition). In: W3C Working Group Note. W3C, 5 Feb 2013. http://www.w3.org/TR/2013/NOTE-rif-primer-20130205/
31. Motik B, Parsia B, Patel-Schneider PF. OWL 2 web ontology language—XML serialization. In: W3C recommendation. W3C, 27 Oct 2009. http://www.w3.org/TR/2009/REC-owl2-xml-serialization-20091027/
32. Motik B, Patel-Schneider PF, Cuenca Grau B. OWL 2 web ontology language—direct semantics (second edition). In: W3C recommendation. W3C, 11 Dec 2012. http://www.w3.org/TR/2012/REC-owl2-direct-semantics-20121211/
33. Motik B, Patel-Schneider PF, Parsia B. OWL 2 web ontology language: structural specification and functional-style syntax (second edition). In: W3C recommendation. W3C, 11 Dec 2012. http://www.w3.org/TR/2012/REC-owl2-syntax-20121211/
34. Ogbuji C. SPARQL 1.1 graph store HTTP protocol. In: W3C recommendation. W3C, 21 Mar 2013. http://www.w3.org/TR/2013/REC-sparql11-http-rdf-update-20130321/
35. Patel-Schneider PF, Motik B. OWL 2 web ontology language—mapping to RDF graphs (second edition). In: W3C recommendation. W3C, 11 Dec 2012. http://www.w3.org/TR/2012/REC-owl2-mapping-to-rdf-20121211/
36. Polleres A, Boley H, Kifer M. RIF datatypes and built-ins 1.0 (second edition). In: W3C recommendation. W3C, 5 Feb 2013. http://www.w3.org/TR/2013/REC-rif-dtb-20130205/
37. Prud'hommeaux E, Buil-Aranda C. SPARQL 1.1 federated query. In: W3C recommendation. W3C, 21 Mar 2013. http://www.w3.org/TR/2013/REC-sparql11-federated-query-20130321/
38. Prud'hommeaux E, Seaborne A. SPARQL query language for RDF. In: W3C recommendation. W3C, 15 Jan 2008. http://www.w3.org/TR/rdf-sparql-query/
39. Schneider M. OWL 2 web ontology language—RDF-based semantics (second edition). In: W3C recommendation. W3C, 11 Dec 2012. http://www.w3.org/TR/2012/REC-owl2-rdf-based-semantics-20121211/
40. Seaborne A. SPARQL 1.1 query results CSV and TSV formats. In: W3C recommendation. W3C, 21 Mar 2013. http://www.w3.org/TR/2013/REC-sparql11-esults-csv-tsv-20130321/
41. Seaborne A. SPARQL 1.1 query results JSON format. In: W3C recommendation. W3C, 21 Mar 2013. http://www.w3.org/TR/2013/REC-sparql11-results-json-20130321/
42. Sporny M, Longley D, Kellogg G, Lanthaler M, Lindstrm N. JSON-LD 1.0—a JSON-based serialization for linked data. In: Sporny M, Kellogg G, Lanthaler M (eds) W3C recommendation. W3C, 16 Jan 2014. http://www.w3.org/TR/2014/REC-json-ld-20140116/
43. Stuckenschmidt H, Schuhmacher M, Knopp J, Meilicke C, Scherp A. On the status of experimental research on the semantic web. In: The semantic web—ISWC 2013: 12th international semantic web conference, lecture notes in computer science, vol 8218, pp 591–606 (2013)
44. ter Horst HJ (2005) Completeness, decidability and complexity of entailment for RDF schema and a semantic extension involving the OWL vocabulary. J Web Semant 3(2–3):79–115. doi:10.1016/j.websem.2005.06.001
45. The W3C SPARQL Working Group. SPARQL 1.1 overview. In: W3C recommendation. W3C, 21 Mar 2013. http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/
46. Tichy W, Lukowicz P, Prechelt L, Heinz E (1995) Experimental evaluation in computer science: a quantitative study. J Syst Softw 28(1):9–18
47. Williams GT. SPARQL 1.1 service description. In: W3C recommendation. W3C, 21 Mar 2013. http://www.w3.org/TR/2013/REC-sparql11-service-description-20130321/
48. Wood D. Whats new in RDF 1.1. In: W3C recommendation. W3C, 25 Feb 2014. http://www.w3.org/TR/rdf11-new/
49. Zimmermann A. RDF 1.1: on semantics of RDF datasets. In: W3C Working Group Note. W3C, 25 Feb 2014. http://www.w3.org/TR/2014/NOTE-rdf11-datasets-20140225/

Birte Glimm is a junior professor at the University of Ulm within the Institute of Artificial Intelligence. Before that, she got her PhD from the University of Manchester in 2007 and worked as research assistant at the University of Oxford until 2011. Her research interests include knowledge representation and reasoning and, in particular, the development and implementation of algorithms and optimizations for reasoning in expressive description logics. She was involved in the standardization of OWL 2 and SPARQL 1.1 and in the organization of several international conferences.

Heiner Stuckenschmidt is a full professor at the University of Mannheim within the Data and Web Science Group. He received his PhD from the Free University of Amsterdam, where he conducted some of the first steps in using The Web Ontology Language for information integration on the Web. Since moving back to Germany in 2005 he leads the AI group at the University of Mannheim. His group is conducting theoretical and applied research on different aspects of semantic web technologies with a focus on semantic matching and combining logical and probabilistic reasoning. More information can be found at http://dws.informatik.uni-mannheim.de.