A grounded ontology for identity and reference of web resources

Aldo Gangemi
Laboratory for Applied Ontology
National Research Council (ISTC-CNR)
Roma, Italy
aldo.gangemi@istc.cnr.it

Valentina Presutti
Laboratory for Applied Ontology
National Research Council (ISTC-CNR)
Roma, Italy
valentina.presutti@istc.cnr.it

ABSTRACT

The identity and reference of web resources used to be a critical subject, and recently it has become a key item in the research agenda of web science. After an introduction to the research related to web identity and reference, we present a new version of IRE, an OWL ontology of web resources and their referencing kinds. Finally, as a case study, we describe an implementation of IRE. Specifically, we show how IRE theory has been used to implement a semantic web application which supports the ontology-driven development and evolution of semantic web portals.

1. INTRODUCTION

The web is an information space realized by computationally accessible resources, each embedding some information, which is encoded in some language, and expresses some meaning. One of the successful achievements of the web is that of allowing different parties of its global communities to share information [21]. Typically, typing an address in a web browser is enough in order to visualize or download an information object, the meaning of which can be then understood by a human agent. Such web address is a URI (Universal Resource Identifier) [5]. There is no doubt about the effectiveness of the URI mechanism for referencing resources that are realized on the web. Nevertheless, there is something more ambitious that the web is supposed to allow than just referencing web resources, that is referencing things in general and allowing people and machines to interoperate, exchange and reuse content.

As underlined by [22]:

> Identification of resources is essential in order to be able to share information about them, reason about them, modify or exchange them.

The semantic web (SW) is an extension of the web which has been developed in order to

> get people to make their data available to others, and to add links to make them accessible by link following. [22]

> ...it will allow data to be shared effectively by wider communities, and to be processed automatically by tools as well as manually. [23]

WWW2007, May 8–12, 2007, Banff, Canada.
.

Based on this vision, we need a simple way to identify resources on the web that can be effective for human agents as well as software agents. We also need to provide web users with easy mechanisms and tools for putting their data on the web in a way that is compliant to the SW vision. We believe that a model of the web could help in undertaking these tasks, and in this paper we make a first step in that direction, by describing a basic OWL [34] ontology which can be grounded in software applications without loosing touch from referencing things in general. As a proof of concept we also refer to a real implementation of our approach named Wikifactory [18], a wiki-oriented tool prototype we have developed that allows that grounding.

Currently, URIs are used as the uniform mechanism for identifying heterogeneous entities, e.g., documents, metadata, ontologies, abstract concepts, physical things, events, multimedia objects. But there is no clear categorization of which are the possible ways to identify and reference those entities on the web.

The issue of identification of resources has been deeply discussed at last "Identity Reference and the Web (IRW2006)" workshop [20], where several interesting approaches have been proposed. In that context, we started the definition of an ontology [16] formalized in first order logic (FOL), which contains a conceptual pattern[1], named "Identity of Resources and Entities" (IRE). IRE has been defined with the aim of addressing the issue of identity, reference, and categorization of resources. Later we presented [39] how this FOL ontology could be encoded in OWL(DL) following two different ontology design pattern-based approaches [13, 33]. In this paper, we present a new version of IRE, and its formalization in OWL(DL). We have expanded IRE's scope in order to model the web primitives. Therefore, we have included concepts for "anchor" elements i.e., elements within web resources that contain attributes with URI as values e.g., HTML `<a href...>`. Our experience of using Semantic Media Wiki [43, 26, 45] and the work on RDFa [1] has inspired us in modeling anchor elements as objects that can be given a semantics.

Furthermore, we discuss how this conceptualization can be used for implementing a certain type of semantic web applications we name *semantic factories*. To this aim we firstly describe the main requirements of a semantic factory (briefly, a semantic web factory provides features for ontology-driven creation, growth, and reuse of semantic web

[1] the term 'pattern' here is used according to the definition of *Content Ontology Design Pattern* given in [13].

portals), and show how IRE has been used to implement a tool i.e., WikiFactory, which addresses such requirements. Specifically, WikiFactory is able: (i) to automatically create a semantic web portal based on a wiki platform starting from an OWL ontology given as input (ii) to automatically generate and update OWL code based on IRE and "Description and Situation (DnS)" theories [15, 41] as a reflection of wiki pages' content (iii) to maintain the synchronization between pages' content and OWL code along the evolution of the semantic web portal, and (iv) to help users in expressing semantics compliant to the underlining ontology.

The rest of the paper is organized as follows: Section 2 briefly discusses related work. In section 3 we introduce IRE, an OWL ontology of resources and their referencing kinds, and in section 4 we present its formalization in OWL. Section 5 describes the requirements of a semantic factory and section 5.1 presents WikiFactory, an example of tool which addresses such requirements by providing an implementation of IRE. Finally, section 6 discusses conclusion and future work.

2. RELATED WORK

According to [22], to identify how to manage (in terms of systems and formalisms) reference and representation of objects on the web, is an important issue of the Web Science. There have been many proposals for addressing these issues. Specifically, some of them approach the problem by providing possible syntactical-oriented solutions [30, 38], some others provide technical-oriented solutions, although approaching the problem also from a conceptual and social perspective [7, 6]. Finally also philosophical/formal approaches have been proposed [10, 35]. For a more detailed discussion of them, see [16, 39].

Another good suggestion comes from Pat Hayes [17] who underlines the difference between *access* and *reference*. Both are relationships between names and things, but they are inherently different and the fact that [21] does not distinguish between the two contributes to cause confusion.

However, none of the above mentioned studies provides a formal model, which could help in finding a comprehensive solution at both syntactic and operational levels. The identity of entities referenced on the web is often intended as the location at which a resource is placed; in other words, there is a need for an explicit distinction between the identity of a resource and its identifier. For example, recalling the W3C web site, the http://www.w3.org URI has its own identity as identifier (i.e., a string), the web location it is associated to has its own identity as a place, the web document has its own identity as a computational object (i.e., file), and the subject of the document has its own identity (i.e., the W3C organization).

Recently, a W3C working group has developed a work on how to embed RDF triples in HTML. There is already a working draft containing a proposal for a syntax, namely RDFa [1]. RDFa complies to our approach although it does not cover all aspects.

3. AN ONTOLOGY OF IDENTITY OF RESOURCES ON THE WEB

The definitions of 'resource' that can be found in literature [21, 4, 5] show ambiguity. This ambiguity makes the issue of handling the identification of a resource very problematic. IRE [16] is our contribution to solve this issue, in this section we present it together with its OWL(DL) formalization. Our approach restricts the nature of the web resource to that of a computational object. This choice is motivated by the fact that a web resource is something that has to be addressable, and things like cars and people are not addressable for their nature. Hence, it is wrong in principle to use the same mechanism of addressing for entities that have such different sorts. IRE has been developed with this principle in mind. As a further principle, we assume that describing the web, as we can observe it in its social and technical behavior, could help addressing the identity issue.

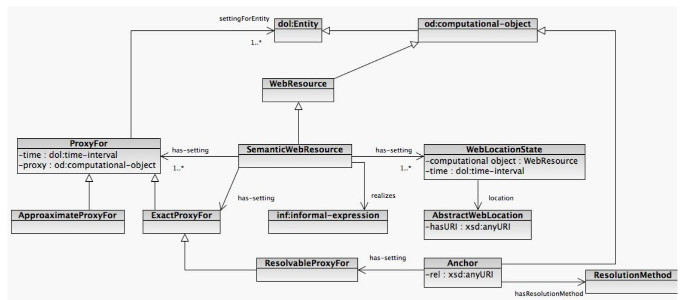

Figure 1: The IRE pattern

Figure 1 depicts IRE. We define *web resources* as computational objects that can be placed in one or more *abstract web locations*. An abstract web location is a place (i.e., a point) in the combinatorial regions that are identified by the URI addressing mechanism (i.e., each URI identifies one and only one abstract web location) [2]. Each abstract web location has only one URI associated with it. Although we can not directly address an entity that is not a computational object, we need to be able to assert facts about it on the web. For this reason we say that a web resource *functions as a proxy for* that entity, at a given time. This means that the identity of web resources is something that goes beyond its location.

In order to make the model clearer, we give a prose description for main elements that have been defined:

WebResource : a computational object that can be composed of other resources. A composed resource has a *partOf* relation with its components. It might have a location (i.e., *WebLocationState*), the address of which is a URI. If the resource is a composed resource the identifier of its abstract location is also an approximate identifier for its parts.

SemanticWebResource : a web resource that realizes an information object through a codification in a formal language for the web (e.g., OWL [34]) which functions as a proxy for whatever entity.

WebLocationState : a situation of a web resource that is associated to a point in the combinatorial regions identified by the URI metric i.e., a web location, at a given time.

[2]Notice that IRIs (Internationalized Resource Identifier) [25] are supposed to replace URIs in next future. Given this, IRIs involvement in the IRE pattern is the same that URIs have.

ProxyFor : a situation of a web resource which functions as a proxy for whatever entity (e.g., a personal home page, a set of metadata describing a person), at a given time.

Anchor : a computational object embedded in a web resource that has an attribute whose value is a URI, and that is associated to a resolution method which causes the dereferencing of the URI e.g., a link between two web resources. An anchor can be given a semantics by means of a semantic relation e.g., OWL datatype or object property.

The IRE pattern of Figure 1 suggests a categorization of web resources that is in their possible role of proxy for an entity. Starting from this particular aspect of a web resource we observed that the class `ProxyFor` can be further specialized. Notice that each specialization of `ProxyFor` can correspond to a different computational approach, or more specifically to a different operational semantic associated to the resolving of the web resource's URI. The kinds of proxy situations can be described informally as follows:

ApproximateProxyFor : is a relationship between a resource and more than one entity at a given time, where the resource is about those entities. In this case the resource represents all the entities approximately.

ExactProxyFor : a relationship between a semantic resource and one entity at a given time, where the semantic resource is about only that entity and describes it through a semantic structure. For example, a set of metadata associated to an individual of an OWL ontology.

ResolvableProxyFor : is a relationship between an anchor and a web resource at a given time, the anchor might allow the access to the web resource it is proxy for.
For example, `<a href="http://www.w3.org">W3C</a>"` in a HTML document is a resolvable proxy for the W3C home page. The anchor might allow to access it by clicking the corresponding link. It may also be have a semantics by means of a semantic relation.

We remark that the relation classes described above are mappable to already existing or proposed concrete solutions. As a proof of concept of this claim, consider the skos property `skos:subjectIndicator` [31] and the topic maps element `subjectIndicatorRef` [28]. The web resource which is the value of such properties would function as a proxy for an entity, this means that the two properties are implementation of either approximate proxies or informal exact proxies[3]. Although this can be a way of identifying the entity of interest the interpretation of the content of the web resource is a responsibility of a human agent. There is no way (at least now) to automatically understand the meaning of the content of a web resource if it is expressed informally. The situation is slightly different if the web resource is a semantic resource. In that case it is expressed with a formal language and functions as an exact proxy for an entity, and it is possible to have a machine able to derive the identity of the entity of interest. For example, this is the case of a set of metadata asserting facts about an individual of a given web ontology that a software agent is able to exploit for performing some inferences.

[3]Notice that the temporal aspect is missing here, however this is compliant with the IRE conceptualization (it is a simplified version of the pattern that leaves time as an implicit variable)

4. IRE PATTERN: FORMAL SPECIFICATION

The IRE pattern just described in prose has been originally specified in first order logic [16], and here we show the OWL ontology [19] corresponding to that formalization, with the extensions described above. The IRE pattern specializes the DOLCE reference ontology [11, 12], and some of its modular extensions, namely Spatial Relations, DnS with Information Objects, and Ontology Design Ontology (ODO) modules. All modules have been developed within EU projects Metokis [29, 3], WonderWeb [27, 14], and NeOn [32]. All modules including IRE ontology are available in OWL at http://www.loa-cnr.it/ontologies/. In the following sections we use a DL syntax notation for expressing axioms according to [36].

The namespace of the IRE ontology [19] is http://www.loa-cnr.it/ontologies/IRE/IRE.owl, hence all ontology elements defined here have that namespace before their specific ID. For the sake of readability we abbreviate the path "`http://www.loa-cnr.it/ontologies`" with "`http://loa/`" while the following prefixes are used in the following definitions that corresponds to the associated namespaces:

$$\begin{aligned}
\texttt{xsd} &= \texttt{http://www.w3.org/2001/XMLSchema\#} \\
\texttt{dol} &= \texttt{http://loa/DOLCE-Lite.owl\#} \\
\texttt{edns} &= \texttt{http://loa/ExtendedDnS.owl\#} \\
\texttt{inf} &= \texttt{http://loa/InformationObjects.owl\#} \\
\texttt{od} &= \texttt{http://loa/OD/OntologyDesign.owl\#}
\end{aligned}$$

4.1 Imported predicates in IRE

The IRE ontology specializes or reuses the following predicates (here we only show the basic taxonomic axioms; for full axiomatization see previously indicated URLs and [27]). From DOLCE [12] IRE imports:

$$\begin{aligned}
&\texttt{dol:Entity, dol:social-object,} \\
&\texttt{dol:region, dol:abstract-region,} \\
&\texttt{dol:time-interval, dol:proper-part-of} \\
\texttt{dol:social-object} &\sqsubseteq \texttt{dol:Entity} \\
\texttt{dol:region} &\sqsubseteq \texttt{dol:Entity} \\
\texttt{dol:abstract-region} &\sqsubseteq \texttt{dol:region} \\
\texttt{dol:time-interval} &\sqsubseteq \texttt{dol:region}
\end{aligned}$$

DOLCE ontology makes basic distinctions between objects, events, and abstract entities. While objects and events like computers and software crashes are spatio-temporally localized, abstract entities like sets and value spaces have no space-time (they are purely formal entities). DOLCE also axiomatizes basic relations such as *part* and *location-of*. From the DnS and Information Objects modules [41] the imports are:

$$\texttt{edns:situation, edns:information-object,} \\ \texttt{edns:information-realization,} \\ \texttt{edns:formal-language, edns:method,} \\ \texttt{edns:realizes, edns:about,} \\ \texttt{edns:ordered-by, edns:involves.}$$

$$\texttt{edns:information-object} \sqsubseteq \texttt{edns:social-object}$$

$$\texttt{edns:information-realization} \equiv \texttt{dol:Entity} \sqcap \\ \exists\, \texttt{edns:realizes.edns:information-object}$$

$$\texttt{edns:method} \sqsubseteq \texttt{edns:social-object}$$

DnS and Information Objects ontologies make basic distinctions between 'descriptive' and 'ground' entities, where the descriptive entities include social objects, like the 'student' or 'professor' roles, the 'being active' task, juridical persons, methods, and also information objects like the text of this paper. Descriptive entities have a lifecycle differently from pure information, which is an abstract entity.

Information objects have a core conceptual pattern, by which they are `edns:ordered-by` a representation language, they are `edns:realized-by` physical information realizations (physical objects, events, etc.), and can be `edns:about` any other entity. Situations are reifications of states of affairs i.e., reifications of n-ary relationships.

From the ODO module [42] IRE imports:

$$\texttt{od:computational-object} \\ \texttt{od:computational-object} \sqsubseteq \\ \texttt{edns:information-realization}$$

The OD ontology specializes the Information Objects ontology in order to build a conceptual schema for digital and analog content to be exchanged during collaborative activities for ontology design. The concept `od:computational-object` specializes `edns:information-realization` for the computational world. Any physical document, electronic service, file, application, etc. are considered here computational objects.

4.2 The IRE predicates and axioms

IRE introduces the following OWL classes, based on predicates described in section 4.1:

$$\texttt{AbstractWebLocation, WebLocationState,} \\ \texttt{WebResource, SemanticWebResource, Anchor,} \\ \texttt{ResolutionMethod, ProxyFor,} \\ \texttt{ExactProxyFor, ResolvableProxyFor,} \\ \texttt{ApproximateProxyFor}$$

the following object properties:

$$\texttt{settingForComputationalObject, settingForTime,} \\ \texttt{settingForWebLocation,} \\ \texttt{settingForProxy, settingForEntity,} \\ \texttt{hasResoulutionMethod}$$

and datatype properties:

$$\texttt{rel, hasURI}$$

The following axioms either characterize or define the above predicates:

$$\texttt{WebResource} \equiv \texttt{od:computational-object} \sqcap \\ \exists\, \texttt{edns:has-setting.ProxyFor} \sqcap \\ \exists\, \texttt{edns:has-setting.WebLocationState} \quad (1)$$

$$\texttt{Anchor} \equiv \texttt{od:computational-object} \sqcap \\ \exists\, \texttt{edns:has-setting.ResolvableProxyFor} \sqcap \\ \exists\, \texttt{hasResolutionMethod.ResolutionMethod} \quad (2)$$

The central concept of IRE is "computational object", a class that is defined in [42]. In IRE classes "web resource" and "anchor" specialize "computational object". We define "web resource" (1) as a computational object that is placed in at least one web location at a given time, and that functions as proxy for some entity, i.e., `dol:Entity`, at a given time. While "anchor" (2) is a computational object which acts as a proxy for only one web resource and furthermore provides a way to access it by means of some resolution method.

These two aspects are formalized by two restrictions on the property `edns:has-setting`, which relates the web resource with situations that, in this case, have to happen. The first is `WebLocationState`, formalized in axiom 3, which means "being in a web location at a certain time", while the second is the `ProxyFor` formalized in axiom 12 which means "being a proxy for some entity at a certain time".

$$\texttt{WebLocationState} \equiv \texttt{edns:situation} \sqcap \\ = 1\ \texttt{settingForComputationalObject}.\top \sqcap \\ = 1\ \texttt{settingForTime}.\top \sqcap \\ = 1\ \texttt{settingForWebLocation}.\top \sqcap \\ \forall\, \texttt{settingForWebLocation.AbstractWebLocation} \sqcap \\ \forall\, \texttt{settingForComputationalObject.WebResource} \sqcap \\ \forall\, \texttt{settingForTime.dol:time-interval} \quad (3)$$

$$\texttt{AbstractWebLocation} \sqsubseteq \texttt{dol:abstract-region} \quad (4)$$

$$\texttt{settingForTime} \sqsubseteq \texttt{edns:setting-for} \sqcap \\ \sqsubseteq \texttt{edns:situation} \times \texttt{dol:time-interval} \quad (5)$$

$$\texttt{settingForWebLocation} \sqsubseteq \texttt{edns:setting-for} \sqcap \\ \sqsubseteq \texttt{WebLocationState} \times \texttt{AbstractWebLocation} \quad (6)$$

$$\texttt{settingForComputationalObject} \sqsubseteq \texttt{edns:setting-for} \sqcap \sqsubseteq \texttt{edns:situation} \times \texttt{od:computational-object} \quad (7)$$

$$\texttt{settingForProxy} \sqsubseteq \texttt{edns:setting-for} \sqcap \sqsubseteq \texttt{ProxyFor} \times \texttt{od:computational-object} \quad (8)$$

$$\texttt{settingForEntity} \sqsubseteq \texttt{edns:setting-for} \sqcap \sqsubseteq \texttt{ProxyFor} \times \texttt{dol:Entity} \quad (9)$$

Object properties defined in definitions from 5 to 9 are subproperties of `edns:setting-for` and are used in relation classes `WebLocationState` and `ProxyFor` in order to specify the intended relationships between web resources and their web location at a given time, and between computational objects and the entity they are proxy for at a given time.

$$\texttt{hasURI} \sqsubseteq \texttt{AbstractWebLocation} \times \texttt{xsd:anyURI} \sqcap \top \sqsubseteq \leq 1\texttt{hasURI}.\top \quad (10)$$

$$\texttt{rel} \sqsubseteq \texttt{Anchor} \times \texttt{xsd:anyURI} \quad (11)$$

Definitions 3 and 12 define two relation classes. This style of design complies with some ontology design patterns [33, 13], and has been chosen because of the temporal nature of the relationships to be modeled (OWL language [34] supports the expression of only binary relations).

Definition 3 defines a relation class that provides a way to relate an abstract web location to a web resource in a given time interval.

A `WebLocationState` is a `dol:situation`, involving exactly three entities: a time interval i.e., `dol:time-interval`, a web resource i.e., `WebResource`, and an abstract web location i.e., `AbstractWebLocation`. Given two separate time intervals, the same abstract web location can be the place of two different web resources
Notice that being a web location of a web resource does not imply the successful resolution of the URI for the abstract web location into the web resource, but only the assignment of an address to the web resource in the combinatorial space identified by URIs.

`AbstractWebLocation`s are DOLCE abstract regions that have only one URI as an identifier. In order to capture the needed semantics we state the axiom 10 that introduces a functional datatype property between `xsd:anyURI` and the OWL class `AbstractWebLocation` that is defined in (4). This design choice applies a design pattern from DOLCE: the possible integration between DOLCE regions and datatypes consists in assuming a datatype structure as a metrics for DOLCE regions. The same patterns applies to `xsd:Date` can be assumed as a metrics for a subset of 'time intervals', which are regions in DOLCE.

Axioms 4, 10 and 3 formalize the "addressing" of computational objects in the virtual web space i.e., web resources.

Definition 3 with related elements provides a model for "referencing" web resources. As we mentioned above, web resources are characterized also by being a proxy for some entity at a given time. In order to model this aspect we define the relation class `ProxyFor` (12).

$$\texttt{ProxyFor} \equiv \texttt{edns:situation} \sqcap = 1\ \texttt{settingForTime}.\top \sqcap = 1\ \texttt{settingForProxy}.\top \sqcap \geq 1\ \texttt{settingForEntity}.\top \sqcap \forall\, \texttt{settingForTime.dol:time-interval} \sqcap \forall\, \texttt{settingForProxy.od:computational-object} \quad (12)$$

Definition 12 formalizes the relation between a computational object, a time interval, and one or more entities. This is a general setting for the *proxy for* relation. `ProxyFor` is specialized for the case of a computational object that is proxy for one and only one entity i.e., `ExactProxyFor` (13), and for the case of a computational object that is proxy for more than one entity i.e., `ApproximateProxyFor` (14). While `ResolvableProxyFor` represents the case anchors.

$$\texttt{ExactProxyFor} \equiv \texttt{ProxyFor} \sqcap = 1\ \texttt{settingForEntity}.\top \quad (13)$$

$$\texttt{ApproximateProxyFor} \equiv \texttt{ProxyFor} \sqcap \geq 2\ \texttt{settingForEntity}.\top \quad (14)$$

$$\texttt{ResolvableProxyFor} \equiv \texttt{ExactProxyFor} \sqcap \forall\, \texttt{settingForProxy.Anchor} \sqcap \forall\, \texttt{settingForEntity.WebResource} \quad (15)$$

Once addressing and referencing has been formalized we can represent access to web resources according to [17]. When a web resource is the result of the resolution of an anchor, we say that it is accessed.

$$\texttt{ResolutionMethod} \sqsubseteq \texttt{edns:method} \quad (16)$$

$$\texttt{hasResolutionMethod} \sqsubseteq \texttt{WebResource} \times \texttt{ResolutionMethod} \quad (17)$$

5. SEMANTIC FACTORIES

We present here an example of IRE application in a community-based SW scenario.
To this aim we introduce the notion of "semantic factory" and explain its requirements. We also present a tool named "WikiFactory", which implements IRE and addresses semantic factory's requirements i.e., WikiFactory is a semantic factory.

The development and evolution of the SW can be affected by *heterogeneity* of expertise and tasks, which can impair semantic interoperability, as well as by the difficulty of associating semantic tags with data and their links. Content on the web is put on by people with any kind of skill, and belonging to different organizations or communities. For example, *web portals* are web sites, often very large, which provide information and services to a community of people, either from a same organization (e.g. an enterprise) or specifically involved (e.g. as experts) in a same domain of interest. When portals exceed a critical size, or when they are linked to (or cover) different domains of expertise, or address different tasks, heterogeneity comes into the ground. Web ontologies are an appropriate infrastructure in order to find local agreements, and to allow applications to reason about data across different domains. But ontologies are difficult to design, map, and validate. An important condition for SW success is to make users able to put and link their data on a SW, without requiring them to be technology experts in semantic technologies. A semantic factory is a SW application, which provides generic web users with ontology-driven features for creating and maintaining SW portals along their whole life-cycle. A semantic factory is a tool which addresses the following requirements:

I: It is based on Web and SW technologies.

II: It supports the automatic creation and growth of SW portals, in an ontology-driven way. This means that the SW portal model consists of an ontology, which is automatically translated to the 'concrete' web pages which compose the SW portal.

III: It is capable to automatically generate and update OWL ontologies starting from web pages and their content. This means that the web pages which compose the SW portal have a correspondent ontological description, which is automatically maintained. Such description is based on some criteria of mapping between the web pages, their content and the ontology/ies.

IV: It automatically maintains the synchronization between SW portal web pages, their content, and the corresponding ontology/ies. This means that according to the mapping criteria, changes in the ontology/ies are reflected to the web pages and their content, and viceversa.

V: It provides user-oriented support to express semantics over the SW, e.g., through suggestions to the users on which relation can be associated with a certain element.

VI: It provides versioning, reuse, and interoperability support.

The notion of semantic factory has been conceived with the aim of defining the main requirements that tools, which support users on the SW should satisfy. Such requirements have been defined in order to address the task of producing 'good' semantic information on the web, which favors interoperability and identification of resources.

5.1 WikiFactory

An example of semantic factory is "WikiFactory". WikiFactory has been designed for SW portals, which run on wiki platforms [24]. Wikis [24] play a leading role among web publishing environments as collaborative tools used for fast and easy input and sharing of content. They are becoming more and more popular, and are particularly suitable in order to create and grow large web portals, e.g. Wikipedia [48]. Semantic wikis are a popular example of SW applications; they are wiki environments where users are allowed to explicitly mark relations between entities and to give them a type in a certain context. A popular example of semantic wikis is Media Wiki (i.e. the one running under Wikipedia) extended by Semantic Media Wiki [43, 26, 45]. Other semantic wikis have been developed so far, such as Semantic Media Wiki, IkeWiki [40], RDFWiki [2], and PeriPeri[9]. The most recent trend is to build a brand new semantic wiki on the basis of new user requirements, and traditional wikis underwent a similar evolution. As a result, there are many available semantic wiki plaforms, each showing its own characteristics and providing different features. The common aspect shared by semantic wikis is the presence of one or more underlying web ontologies, or more generally the use of SW technologies.

Besides these basic technical features, we may wonder if semantic wikis could also allow users to express an "inherently validated" semantics. Do members of a community actually collaborate (more or less consciously) towards the creation and maintainance of an appropriate web ontology? For example, when a user tags a text by using classes and relations which were previously created by himself/herself or by other users from the same community? With respect to that, some limitations seem to apply. While semantic wikis appear extremely useful, and contribute to spread the use of web ontologies among generic web users, they might find it difficult, for example, to choose which relation should be used in a given context, especially if they want to be compliant to a suitable ontology. Additionally, they may not want to be in charge of preserving formal consistency, nor to be assigned time-consuming and possibly boring tasks, such as learning and using a new, possibly complex, syntax, or knowing all appropriate categories/relations in advance. WikiFactory[4] [18] has been designed with these rationales in mind. It is a semantic factory, which implements IRE so as to provide a proof of its practical usefulness. WikiFactory is already available and downloadable at [46], although in "alpha" version for testing: new versions are uploaded weekly. We remark that WikiFactory has not to be considered a main contribution of this work because we still need additional test results for a satisfying evaluation of qualitative and quantitative parameters. However, we think it is worth to briefly describe it, in order to provide a proof of concept of our proposal, and to make it more understandable. Referring to the definitions of semantic factory's requirements above given, WikiFactory is characterized by the followings:

I. It is implemented in Java with web services support, it works with OWL ontologies, and performs the deployment in an XML-based format called Wiki Interchange Format (WIF) [44], which has been extended in order to support RDFa features. The use of WIF format provides WikiFactory with the property of being platform-independent with regard to semantic wiki platforms.

II. It takes in input an OWL ontology describing a specific domain of interest, possibly together with its individuals. It is able to automatically deploy a SW portal on a semantic wiki platform, starting from such OWL description. Current implementation of WikiFactory associates each element of the ontology with a wiki page[5]. We have chosen to perform the deployment as an extended WIF serializa-

[4] WikiFactory is being developed with the effort of a team made of students, junior, and senior researchers. The reader may visit the WikiFactory homepage at http://semanticfactory.org/wikifactory for more detail

[5] Next version of the tool will support page 'fragments': each ontology element will be managed as a page fragment.

tion format, in order to make the tool usable with any semantic wiki platform, by simply providing an appropriate “WIF to Wiki” translator. The current version of WikiFactory provides a translator to Semantic Media Wiki. Notice that our extended WIF format is more expressive than any available wiki, since it is a superset of the set containing all currently supported constructs. It is up to the developer of the translator to decide what to support and what to ignore.WikiFactory deployment functionality can be compared to other model-driven applications, e.g. WebML, a tool for creating web applications [8].

III. Each wiki page has its correspondent elements in the OWL ontology. WikiFactory distinguishes between content and structure. It automatically generates two separated OWL ontologies: one representing the web portal structure, the other representing the domain of knowledge i.e., it deals with pages' content.

IV. WikiFactory monitors content changes of the web portal and reflects them on the underlying ontology, and viceversa. Therefore, the domain ontology and the web portal evolve together, and are mutually representative of each other.

V. In the requirement analysis of WikiFactory we have played down functionalities for consistency checking on the growing ontology. It is intuitive that the more constructs we want to support, the less performing the tool would result. However, WikiFactory is interfaced to a DL reasoner (currently, Pellet [37]), so that checking, realization, and classification are just one step away, if needed. Nonetheless, our choice has been to “give guidance” to users, in order to let them express semantics as safely as possible. On one hand user-oriented support allows users to express any OWL constructs, on the other hand it contextually disallows those constructs that can result as mistakes. For example, the following operations are automatically carried out:

- To disallow the semantic relation between two wiki pages a and b with a property p when the domain of p is disjoint from the class of a [6];
- To put a placeholder for property p in all pages representing individuals whose class has an existential restriction on p;
- In each wiki page, all properties applicable to that individual are listed[7]
- The range of values that can be expected for a property are suggested based on universal restrictions on that property;

These are some examples of functionalities that can be implemented in order to help the user in expressing a domain-oriented semantics. We have only listed those that are already supported by WikiFactory. We plan to implement many more features by using a mix of statistical, NLP and ontology engineering techniques.

VI. WikiFactory maintains versions of the grounded ontology. Those versions can then be used for several purposes: to deploy a new web portal by using a specific version of the ontology, to reuse the ontology in other applications of the same domain or organization, to study ontology evolution, etc.

The mapping criteria which **II., III,** and **IV.** relies on IRE and DnS ontologies. There are two main aspects of WikiFactory that can be observed as direct implementation of IRE. The first aspect is that it automatically generates an OWL ontology describing the web portal[8]. Such ontology distinguishes between description of wiki pages (i.e., as web resources) and domain elements (i.e., as entities) and links them by means of instances of the IRE `ProxyFor` relation class (defined in section 4). Figure 2 shows how wiki pages are described in terms of IRE. A wiki page is always an exact proxy for any entity, i.e. an instance of any class from a domain ontology. Domain concepts are reifications of classes defined in the domain ontology[9]. This definition of wiki pages is inspired by the intended meaning of today's usage of wikis running on the Semantic Media Wiki platform [43, 26, 45]. However, we plan to manage “fragments” i.e., pieces of wiki pages, in the same way wiki pages are currently handled in WikiFactory. The second aspect is that WikiFactory

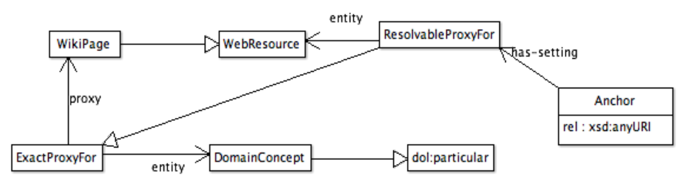

Figure 2: Wiki basic elements in terms of IRE

allows users to express semantic relations embedded within anchors. Two cases are possible:

- an anchor, which embeds a semantic relation, connects two internal wikipages. This results in a behavior akin to that of Semantic Media Wiki,
- an anchor, which embeds a semantic relation, connects an internal wiki page to a web resource, which is external to the wiki. This is reported in the ontology as well. Notice that when we semantically connect a wiki page (internal to a specific web portal) to some external web resource by means of a guided system like this, we are “saying” something about what the external web resource describes. Furthermore, we are creating semantic relations across different domains which could possibly be considered to some extent reliable (at least within the web portal community expressing such links). This feature could be used e.g. to express semantic relations that map concepts across different web portals, which support related domains.

6. CONCLUSION AND FUTURE WORK

[6]This semantics is motivated by the fact that the subject and object of the triple expressing the relationship are two individuals having the two wikipages as proxies

[7]Actually, only the properties that explicitly have the class of that individual as their domain are included by default

[8]Actually at least two separated ontologies or ontology modules are generated. For the sake of simplicity we refer to a single 'union' ontology

[9]This reflects the application of the *Description and Situation* theory, which has been also used in the design of IRE, see section 4

As underlined by [22] identification of resources is an important task to use them on the web. We think that Semantic Web currently provides the most suitable mechanisms and technologies for studying solutions to this problems. Nevertheless, in order to exploit Semantic Web solutions, there is a strong need for tools which enable generic web users to produce semantic information on the web, which is "good" for interoperability objectives. In this paper we have presented IRE, an OWL ontology of web resource identity and reference. Following [17], IRE provides a model for "referencing" web resources, for "addressing" computational objects in the virtual web space, and for "accessing" web resources. The core concepts of IRE are web resource" and "anchor". The first is formalized as a computational object that has a web location at a given time, and functions as a proxy for some entity. The second (a typical example of anchor is the HTML "a" tag) is formalized as a computational object, which is embedded in web resources and may allow the access to a web resource, and that can be associated to semantic relations. In order to provide a proof of concept of the practical usefulness of our proposal, we have introduce the notion of "semantic factory". A semantic factory is a type of Semantic Web tools, which support web users in using and contributing to the Semantic Web. Specifically, we have described a specific semantic factory named "WikiFactory" [47, 46, 18] which is able to automatically create and maintain a Semantic Web portal starting from an OWL ontology. What is more, WikiFactory automatically maintains the synchronization between the semantic web portal structure, its content, and the OWL ontology. This synchronization relies on a mapping theory which is based on IRE and *Description and Situations* ontologies [15]. We have also shown how WikiFactory implements features that "give guidance" to the users in order to let them express semantics as safely as possible.

Finally, we remark the potential benefits of the automatic generation of "semantic anchors", e.g. when they are resolved out from the semantic web portal, they carry a proper semantics, and can 'transfer' that semantics to external web resources, thus creating de facto semantic relations across domains. A possible evolution of our work is the study of mechanisms for managing agreement on collaborative ontology creation and maintenance with semantic web factories. A special case is ontology mapping, since it can be represented by semantic anchors that link pages of different wikis. We are also considering the use of Natural Language Processing techniques to assist the user in choosing/creating new semantic information as well as to learn semantic links from anchors.

7. ACKNOWLEDGMENTS

We are grateful to the members of the NeOn consortium who contributed to the NeOn vision being funded by the European Commission 6th IST Framework Programme. Further information on NeOn is available on http://www.neon-project.org.

8. REFERENCES

[1] B. Adida and M. Birbeck. RDFa Primer 1.0 Embedding RDF in XHTML. Technical report, World Wide Web Consortium, May 2006. http://www.w3.org/TR/xhtml-rdfa-primer/.

[2] P. S. B. RDFWiki. http://infomesh.net/2001/rdfwiki/.

[3] W. Behrendt, A. Gangemi, W. Maass, and R. Westenthaler. Towards an Ontology-Based Distributed Architecture for Paid Content. In *The Semantic Web: Research and Applications: Second European Semantic Web Conference, ESWC, May 29 - June 1, 2005. Proceedings*, volume 3532, pages 257–271, Heraklion, Crete, Greece, 2005. Springer-Verlag GmbH.

[4] T. Berners-Lee, R. T. Fielding, and L. Masinter. Uniform Resource Identifier (URI): Generic syntax. Internet Draft standard RFC2396, The Internet Society, August 1998. http://rfc.net/rfc2396.html.

[5] T. Berners-Lee, R. T. Fielding, and L. Masinter. Uniform Resource Identifier (URI): Generic syntax. Internet Draft standard RFC3986, The Internet Society, January 2005. http://rfc.net/rfc3986.html.

[6] J. Black. Creating a Common Ground for URI Meaning Using Socially Constructed Web sites. Architecture and Philosophy of the Web Identity, Reference, and the Web (IRW2006), WWW2006 Workshop, Edinburgh, Scotland, May, 23rd 2006. http://www.ibiblio.org/hhalpin/irw2006/jblack.pdf.

[7] D. Booth. Four Uses of a URL: Name, Concept, Web Location and Document Instance. http://www.w3.org/2002/11/dboothnames/dbooth-names_clean.htm, 2003.

[8] S. Ceri, F. Daniel, M. Matera, and F. Facca. Model-driven Development of Context-Aware Web Applications. *To appear on ACM Transactions on Internet Technology (ACM TOIT)*, 7, May 2007.

[9] Chris Purcell. Periperi. http://www.srcf.ucam.org/ cjp39/Peri/PeriPeri.

[10] K. Clark. Identity Crisis. http://www.xml.com/pub/a/2002/09/11/deviant.html, 2002.

[11] DOLCE - Project Home Page. http://dolce.semanticweb.org.

[12] DOLCE Ontology. http://www.loa-cnr.it/ontologies/DOLCE-Lite.owl.

[13] A. Gangemi. Ontology Design Patterns for Semantic Web Content. In *M. Musen et al. (eds.): Proceedings of the Fourth International Semantic Web Conference*, Galway, Ireland, 2005. Springer.

[14] A. Gangemi, S. Borgo, C. Catenacci, and J. Lehmann. Task Taxonomies for Knowledge Content. Deliverable D07 of the Metokis Project. http://www.loa-cnr.it/Papers/D07_v21a.pdf, 2005.

[15] A. Gangemi and P. Mika. Understanding the Semantic Web through Descriptions and Situations. In *Meersman R. et al. (eds.) - Proceedings of ODBASE03 Conference*. Springer, 2003.

[16] A. Gangemi and V. Presutti. The bourne identity of a web resource. Architecture and Philosophy of the Web Identity, Reference, and the Web (IRW2006), WWW2006 Workshop, Edinburgh, Scotland, May, 23rd 2006. http://www.ibiblio.org/hhalpin/irw2006/presentations/-vpresutti.pdf.

[17] P. Hayes. In defense of Ambiguity. Architecture and

Philosophy of the Web Identity, Reference, and the Web (IRW2006), WWW2006 Workshop, Edinburgh, Scotland, May, 23rd 2006. http://www.ibiblio.org/hhalpin/irw2006/presentations/-HayesSlides.pdf.

[18] A. D. Iorio, V. Presutti, and F. Vitali. Wikifactory: An ontology-based application for creating domain-oriented wikis. In Y. Sure and J. Domingue, editors, *ESWC*, volume 4011 of *Lecture Notes in Computer Science*, pages 664–678. Springer, 2006.

[19] Identity of resource and entity on the web ontology. http://wiki.loa-cnr.it/index.php/LoaWiki:IRE.

[20] Identity reference and the web, www2006 workshop. http://www.ibiblio.org/hhalpin/irw2006/.

[21] I. Jacobs and N. Walsh. Architecture of the World Wide Web. Technical report, W3C, 2004. http://www.w3.org/TR/webarch.

[22] T. B. Lee, W. Hall, J. A. Hendler, K. O'Hara, N. Shadbolt, and D. J. Weitzner. A Framework for Web Science. *Foundations and Trends in Web Science*, 1:1–130, 2006.

[23] T. B. Lee, W. Hall, J. A. Hendler, N. Shadbolt, and D. J. Weitzner. Web Science. *Science*, 313, August 2006.

[24] B. Leuf and W. Cunningham. *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley, 2001.

[25] D. M. and S. M. Internationalized Resource Identifiers (IRIs). Internet Draft standard RFC3987, The Internet Society, January 2005. http://www.ietf.org/rfc/rfc3987.txt.

[26] K. M., V. D., and V. M. Wikipedia and the Semantic Web: The Missing Links. In *Proceedings of Wikimania 2005*, Frankfurt, Germany, August 2005.

[27] C. Masolo, A. Gangemi, N. Guarino, A. Oltramari, and L. Schneider. Wonderweb eu project deliverable d18: The wonderweb library of foundational ontologies. http://wonderweb.semanticweb.org/deliverables/-documents/D18.pdf, 2004.

[28] Members of the Topic Maps.Org Authoring Group. Xml topic maps (xtm 1.0). Technical report, TopicMaps.org, 2001. http://www.topicmaps.org/xtm/1.0/.

[29] Metokis Project. http://metokis.salzburgresearch.at.

[30] A. Miles. Working Around the Identity Crisis. http://esw.w3.org/topic/SkosDev/IdentityCrisis, 2005.

[31] A. Miles and D. Brickley. SKOS Core Vocabulary Specification. Technical report, World Wide Web Consortium (W3C), November 2005. http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102/.

[32] NeOn EU project home page. http://neon-project.org.

[33] Noy N. and Rector A. Defining N-ary Relations on the Semantic Web: Use With Individuals. Technical report, W3C, 2005. http://www.w3.org/TR/swbp-n-aryRelations/ (2004).

[34] OWL Web Ontology Language Family of Specifications. http://www.w3.org/2004/OWL, 2004.

[35] B. Parsia and P. F. Patel-Schneider. Meaning and the Semantic Web. Architecture and Philosophy of the Web Identity, Reference, and the Web (IRW2006), WWW2006 Workshop, Edinburgh, Scotland, May, 23rd 2006. http://www.ibiblio.org/hhalpin/irw2006/bparsia.pdf.

[36] P. Patel-Schneider. Working Version (Draft): Description-Logic Knowledge Representation System Specification from the KRSS Group of the ARPA Knowledge Sharing Effort.

[37] Pellet home page. http://pellet.owldl.com/.

[38] S. Pepper and S. Schwab. Curing the Web's Identity Crisis: Subject Indicators for RDF. Technical report, Ontopia, 2003. http://www.ontopia.net/topicmaps/-materials/identitycrisis.html (2003).

[39] V. Presutti and A. Gangemi. Towards an OWL Ontology for Identity on the Web. In *In Proceedings of SWAP 2006, the 3rd Italian Semantic Web Workshop*, Pisa, Italy, December 18-20 2006. CEUR Workshop Proceedings.

[40] S. Schaffert, D. Bischof, T. Brger, A. Gruber, W. Hilzensauer, and S. Schaffert. Learning with semantic wikis. In M. Völkel and S. Schaffert, editors, *Proceedings of the First Workshop on Semantic Wikis – From Wiki To Semantics*, Workshop on Semantic Wikis. ESWC2006, June 2006.

[41] The Description and Situation ontology. http://www.loa-cnr.it/ontologies/ExtendedDnS.owl.

[42] The Ontology Design ontology. http://www.loa-cnr.it/ontologies/OD/OntologyDesign.owl.

[43] M. Völkel, M. Krötzsch, D. Vrandecic, H. Haller, and R. Studer. Semantic Wikipedia. In *Proceedings of the 15th International World Wide Web Conference (WWW 2006)*, Edinburgh, 2006.

[44] M. Vlkel and E. Oren. Towards a wiki interchange format (wif). In *First Workshop on Semantic Wikis – From Wiki To Semantics, Budva, Montenegro, 12.06.06*, 2006.

[45] S. M. Wiki. Semantic MediaWiki Home Page. http://meta.wikimedia.org/wiki/Semantic_MediaWiki.

[46] WikiFactory download page. http://wiki.loa-cnr.it/index.php/LoaWiki:Project:WikiFactoryDownloads.

[47] WikiFactory download page. http://wiki.loa-cnr.it/index.php/LoaWiki:Project:WikiFactory.

[48] Wikipedia. Wikipedia Home Page. http://www.wikipedia.org.