The bourne identity of a web resource

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

One of the main strength of the web is that it realizes the goal of allowing any party of its global community to share information with any other party. This has been achieved by making use of a unique and uniform mechanism of identification, the URI (Universal Resource Identifiers). Web applications such as search engines have been built up on this mechanism. Although URIs succeed when used for retrieving resources on the web, their suitability as a way for identifying any kind of *thing*, for example resources which are not on the web, is not guaranteed.

We investigate the meaning of identity of a resource and how it can be modeled in order to be implemented on the web. In particular, we use an ontological approach for modeling the problem, and discuss some possible solutions. We describe the concept of resource from the web domain point of view, using an ontology of Information Objects, built under DOLCE and some of its modular extensions. In particular, we formalize some concepts of *resource* and distinguish them from the concept of a generic *entity*. We finally propose a formal pattern for the categorization of resources.

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 [10]. Typically, typing an address in a web browser is enough in order to visualize or download an object, the meaning of which can be then understood by a human agent. Such web address is a URI (Universal Resource Identifier) [3]. There is no doubt about the effectiveness of the URI mechanism for the referencing of 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. That ambitious goal requires a software agent to identify a resource in an unambiguous way, in order to perform the appropriate operations on it.

Currently, URIs are used as the uniform mechanism for identifying heterogeneous entities, e.g., documents, metadata, ontologies, abstract concepts, physical things, events. But there is no clear categorization of which are the possible ways to identify and reference those entities on the web. This sort of confusion has led to lack of consensus on which is the most suitable way to solve the problem of handling the recognition of the identity of an element that is referenced by a URI, and consequently there is not a defined operational semantics associated with each of these different sorts. We propose a formal description of the main concepts that compose the architecture of the web, by means of an ontological approach. To this aim, we introduce IRE (Identifiers, Resources, and Entities), a conceptual pattern based on an ontology of Information Objects [1], built on top of DOLCE and on some of its modular extensions [15, 9]. For a complete report the reader can refer to [8, 14], while all OWL files are available at http://www.loa-cnr.it/ontologies. IRE provides the basis for defining a categorization of the kinds of resources that can be referenced on the web. We are confident that, based on this categorization, it is possible to study the most suitable way of handling the operational semantics that can be applied to different references.

The rest of the paper is organized as follows: in section 1.1 we discuss briefly the existing approaches to the problem of identifying a resource on the web, in section 2 we discuss about the definitions of the concept of resource and its relation to the URI mechanism of identification that can be found in normative documents such as [3] and [10]. In particular, we underline the ambiguity of such definitions, showing the need for a rationale. Finally, we present the IRE pattern, the most specific contribution of this paper. A general description is given in section 3 and in section 4 we report its formal specification. The section 5 summarizes the proposal underlining our contribution.

1.1 Related work

The identification of resources is an important task to use them on the web. Currently, there is a diffuse feeling that the identification procedures suffer from a lack of consensus about how to handle them. This feeling is sometimes referred to as 'identity crisis' on the web. A typical example is the URI http://www.w3.org: does it identify the web document that is placed at that web address, or the W3C organization? There have been many proposals suggesting different approaches to the aim of addressing the issue. In this section we briefly summarize some significant ones.

Alistar Miles [18] describes his perception of the problem by isolating a possible obstacle: the creation of a same URI for representing different concepts. This has also been named *URI collision* [10]. Miles proposes an interesting 'low level' approach as a best practice, that of using HTTP URIs to address entities that are not accessible on the web. He proposes to manage the problem at the server side by

WWW2006, May 22–26, 2006, Edinburgh, UK.
.

means of a negotiation on how to resolve the URI. For example, if one creates the URI http://foo.com/me to describe herself, then it could be resolved by the server as the URI http://foo.com/me.html or http://foo.com/me.xml or other, depending on a sort of configuration of the browser.

Steve Pepper [21] expresses a similar difficulty about the use of URIs for identifying all kinds of entities. In particular, he proposes to associate a resource to a document, whose content describes the subject of the resource (i.e., a subject indicator). Nevertheless, this solution leaves the responsibility of interpreting the identity of a resource to a human agent, and there is no way to ensure that the subject indicator refers to a single subject.

Kendall Clark [5] discusses the 'tidiness' of web specifications, and the importance to clarify the conceptual assumptions upon which the web is built, and the semantic web is going to be built.

David Booth [4] proposes an informal categorization of what can be identified by a URI, suggesting the definition of different conventions for each of the four uses he has identified.

All these proposals are important contributions to solve the identity crisis but no one of them is supported by a provided formal model, which could help in finding a comprehensive solution also at both a syntactic and operational level. Furthermore, the identity of entities referenced on the web is often intended as the location in which a resource is placed; in other words, there would be needed 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), the subject of the document has its own identity (i.e., the W3C organization).

2. URI AND RESOURCES

The term "resource" is generally used for all things that might be identified by a URI [10]. In literature, we find several definitions for the term "resource" used in the context of world wide web. In particular we quote here three authoritative documents, [3, 2, 10][1] and discuss about the way and consequences of the definition they provide for "resource" . In [2] the concept of resource is defined as follows:

> A resource can be anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., "today's weather report for Los Angeles"), and a collection of other resources. Not all resources are network "retrievable"; e.g., human beings, corporations, and bound books in a library can also be considered resources. The resource is the conceptual mapping to an entity or set of entities, not necessarily the entity which corresponds to that mapping at any particular instance in time. Thus, a resource can remain constant even when its content—the entities to which it currently corresponds—changes over time, provided that the conceptual mapping is not changed in the process.

[1]Note that [2] has been replaced by [3] but we decided to quote and discuss about it here because it is still used as reference from many web users

Moreover, in the same document the mechanism for identifying resources (i.e., URI) is specified, and it is also said that:

> Having identified a resource, a system may perform a variety of operations on the resource, as might be characterized by such words as 'access', 'update', 'replace', or 'find attributes'.

The following definition of "resource" is given by [3] which updates [2]:

> This specification does not limit the scope of what might be a resource; rather, the term "resource" is used in a general sense for whatever might be identified by a URI. Familiar examples include an electronic document, an image, a source of information with a consistent purpose (e.g., "today's weather report for Los Angeles"), a service (e.g., an HTTP-to-SMS gateway), and a collection of other resources. A resource is not necessarily accessible via the Internet; e.g., human beings, corporations, and bound books in a library can also be resources. Likewise, abstract concepts can be resources, such as the operators and operands of a mathematical equation, the types of a relationship (e.g., "parent" or "employee"), or numeric values (e.g., zero, one, and infinity).

On the other hand, in [10] the concept of resource is used with twofold meaning: that of whatever might be identified by a URI, and that of anything that can be the subject of a discourse, such as cars, people and so on. Furthermore, the concept of *Information resources* is defined as those resources which essential characteristics can be conveyed in a message. [10] also defines the principle of *opacity* of a URI, which promotes the independence between an identifier and the state of the identified resource.

Given the above, at least foour possible interpretations of the intended meaning of the term "resource" emerge. Even though it is not our principal aim, it could be useful to establish what meaning is the most suitable in the web domain.

- **computational object:** if a resource is a computational object, something on which one can perform operations [2] - in this context we define "computational object" such as (i) the physical realization of an information object, (ii) something that can participate in a computational process. Examples of computational objects are: a database, a digital document, a software application - then its identity would not be equivalent to a virtual localization, because a computational object is a physical entity and realizes (is the support for) a certain information object. Neither physical entities nor information objects can be reduced to regions in a virtual space, especially if that space should be uniquely identifiable through URIs. For example, the personal home page of Aldo Gangemi is a document which exists on the web and is reachable through the dereferencing of its URI, but it does continue to exist also if it changes its location or if the server it is stored on becomes offline.

- **conceptual mapping:** if a resource is intended as a "conceptual mapping" [2] then its identity is purely formal. For this reason it cannot be also intended as a “computational object”. As a conceptual mapping, a resource can be characterized as a location in the virtual space of the combinatorial regions that are identified by the URIs. Consequently, the identity of a resource in this sense is equivalent to a localization in that space. As a matter of fact, without that space, it would not exist, and its URI is sufficient to identify it unambiguously.

- **proxy:** considering the principle of *opacity* [10], the sense of resource can be that of a ’proxy’ that is localized in a region of the virtual space identified by the URIs. In this case, the resource is actually intended as a computational object, and its identity is given by the set of elements composing the proxy. For example, an English text, a picture, a metadata schema, etc. According to this meaning of “resource”, its identity goes beyond its location. A resource does exist beyond its location, and its identity holds over its presence on the web.

- **entity:** by defining “resource” with the meaning of an entity [3] - being it either a computational object or not - is problematic because the relationship that holds between a resource and a URI would be the same for addressing computational objects and physical or abstract objects. This approach is problematic, because it attempts to address entities (i.e., physical and abstract objects) that are not addressable in principle.

Based on this rationale, we give a formal description of the different meanings associated to the term ’resource’, and contextualize those meanings in a formal pattern. It has to be noticed that the concept associated with the term “resource” here is different from that associated to the same term in [3]. The definition in [3] corresponds to that of concept “entity” used in IRE. We reserve the name ’resource’ to one meaning, which seems more intuitive from a common-sense viewpoint, even though our principal aim is conceptual, not terminological.

3. THE IRE PATTERN

As outlined above, the definitions of resource that can be found in literature [10, 2, 3] show ambiguity, making the issue of handling the identification of a resource very problematic.

Our approach restricts the nature of the resource to that of a computational object. This choice is motivated by the fact that a 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.

The Figure 1 depicts the IRE (Identifiers, Resources, and Entities) pattern that we propose for defining the main concepts that are useful for approaching the issues of defining a categorization of resources and the way of identifying them.

We define a *Resource* as a computational object that has a particular feature, it can be placed in one or more *AbstractWebLocation*. This means that the identity of the resource is something that goes beyond its location. An abstract web location is a place (i.e., a point) in the combinatorial regions that are identified by the URI addressing

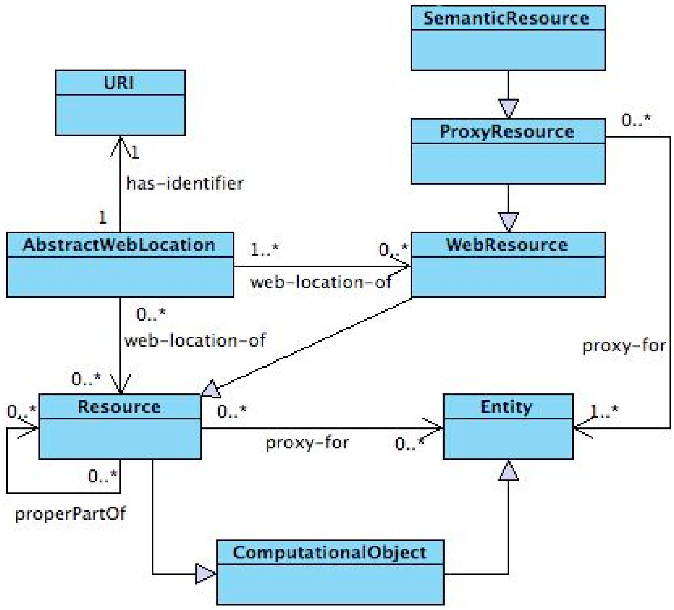

Figure 1: The IRE pattern

mechanism (i.e., each URI identifies one and only one abstract web location) [2]. To this aim the datatype property *has-identifier* is specified as functional. If a resource is made available on the world wide web then it is a *WebResource*. In this sense to be a web resource is a particular state of a resource. 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. We can do this by using a web resource whose functions as a proxy for that entity (i.e., a *ProxyResource*). In order to make the model clearer, we give a prose description for each element that has been defined:

AbstractWebLocation : a point in the combinatorial regions identified by the URI metric.

Resource : 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.,*AbstractWebLocation*), 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.

WebResource : a resource that is made available on the world wide web, hence accessible through a web protocol (e.g., a document, a web service).

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

SemanticResource : a web resource that realizes an information object through a codification in a formal language for the web (e.g., OWL [13]) which functions as a proxy for whatever entity.

[2]Notice that IRIs (Internationalized Resource Identifier) [11] are supposed to replace URIs in next future. Given this, IRIs involvement in the IRE pattern is the same that URIs have.

The IRE pattern of Figure 1 suggests a categorization of the resources that is in their possible role of functioning as a proxy for an entity. Starting from this particular feature of a resource we noticed that the relation *proxyFor* can be of four types, and that each of them can be treated in a different way. In fact, each kind of instance of the *proxyFor* relation can correspond to a different computational approach, or more specifically to a different operational semantic associated to the resolving of the proxy's URI. The four kinds of proxy relations can be described informally as follows:

- resolvable proxy for: is a relationship between a proxy resource and a web resource where the proxy resource allows the access to the web resource it is proxy for. For example, a 'href' in a HTML document is a resolvable proxy for the web resource it allows to access by clicking the corresponding link.
- approximate proxy for: is a relationship between a resource and more than one entity, where the resource is about those entities. In this case the resource represents all the entities approximately.
- informal exact proxy for: a relationship between a resource and one entity, where the resource is about only that entity. For example a document which subject of discourse is exactly the entity that the resource is proxy for.
- formal exact proxy for: a relationship between a semantic resource and one entity, 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.

We remark that the relations 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` [19] and the topic maps element `subjectIndicatorRef` [16]. The web resource which is the value of such properties would function as a proxy resource for an entity, this means that the two properties are either approximate proxies or informal exact proxies. 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 functions as a formal 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.

3.1 An example: IRE pattern and Topic Maps

In order to give an example of the applicability of this approach also to techniques other than RDF, we informally describe how to classify the Topic Maps concepts related to identity issues in terms of IRE. Based on [12]:

> "..subject - anything whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever..".

The Topic Maps concept of "subject" is equivalent to that of "Entity". More precisely, the terms "subject" and "entity" are synonyms. From the same document [12] we have the following definition of topic:

> A topic is a symbol used within a topic map to represent one, and only one, subject, in order to allow statements to be made about the subject through the assignment of characteristics to the topic.

Based on IRE pattern a topic in a topic map is a *formal exact proxy* (i.e., a semantic resource) for a given entity. Again from [12] we have the following definitions about identification of entities:

> "A subject indicator is an information resource that is referred to from a topic map in an attempt to unambiguously identify the subject of a topic to a human being. Any information resource can become a subject indicator by being referred to as such from within some topic map, whether or not it was intended by its publisher to be a subject indicator. A subject identifier is a locator that refers to a subject indicator. Topic maps contain only subject identifiers (and not the corresponding subject indicators)...." "...A subject locator is a locator that refers to the information resource that is the subject of a topic. The topic thus represents that particular information resource; i.e. the information resource is the subject of the topic... "

In terms of IRE pattern, a "subject indicator" is a web resource which acts as *approximate proxy for* a given entity. In order to identify an entity by a subject indicator, a human interpretation is needed, and it is not guaranteed that such identification is unambiguous. Furthermore, a subject identifier is an abstract web location (i.e. a URI), which locates a subject indicator on the Web. On the other hand, when the entity is addressed by a "subject locator", it can be considered a computational object available on the web, i.e., a web resource.

4. IRE PATTERN: FORMAL SPECIFICATION

The IRE pattern just described in prose has been specified in first order logic, such formalization is presented in this section. In particular, in section 4.1 some reused and specialized predicates are presented and the section 4.2 shows the IRE model theory.

4.1 Imported predicates in IRE

The IRE (Identifiers, Resources, and Entities) pattern specializes the DOLCE reference ontology [6, 7], and some of its modular extensions, namely Spatial Relations, DnS with Information Objects, and Knowledge Content Objects (KCO) modules (the last has been developed in the EU Metokis project [17, 1]. All modules are available in OWL at http://www.loa-cnr.it/ontologies/. For a complete report the reader can refer to [14, 8].

The IRE ontology specializes or reuses the following predicates (classes have capitalized names, here we only show the basic taxonomic axioms; for full axiomatization see indicated URLs and [14]). From DOLCE [6] IRE imports:

$$\{Entity, SocialObject, Region, AbstractRegion, TimeInterval, properPartOf\}$$

$$SocialObject(x) \rightarrow Entity(x)$$
$$Region(x) \rightarrow Entity(x)$$
$$AbstractRegion(x) \rightarrow Region(x)$$
$$TimeInterval(x) \rightarrow Region(x)$$

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 abstract value spaces have no space-time (they are purely formal entities). DOLCE also axiomatizes basic relations such as *part* and $locationOf$.
From the DnS and Information Objects modules (http://www.loa-cnr.it/ontologies/ExtendedDnS.owl) the imports are:

$$\{InformationObject, InformationRealization, FormalLanguage, Method, realizes, about, orderedBy, involves\}$$

$$InformationObject(x) \rightarrow SocialObject(x)$$
$$InformationRealization(x) \rightarrow Entity(x)$$
$$Method(x) \rightarrow SocialObject(x)$$

DnS and Information Objects ontology makes basic distinctions between 'descriptive' and 'ground' entities, where the descriptive part of a conceptualization is only made of social objects, like the 'student' or 'professor' roles, the 'being active' task, juridical persons, formal languages rules, methods, and including also information objects like the pure text of this paper. Descriptive entities have a lifecycle (differently from pure information, which is a formal entity).
Information objects have a core conceptual pattern, by which they are *orderedBy* a representation language, they are *realizedby* physical information realizations (physical objects, events, etc.), and can be *about* any other entity.
From the Spatial Relations module (http://www.loa-cnr.it/ontologies/SpatialRelations.owl) the import is:

$$\{eAbstractLocationOf\}$$

The Spatial Relations ontology takes some basic primitives and axioms from e.g. spatial reasoning theories, and generalizes them to talk about 'absolute' locations (e.g. the location of a ship at geographic coordinates) and 'relative' locations (e.g. the location of a ship with reference to a coast). The relation $eAbstractLocationOf$ relation, imported in IRE, holds between abstract regions (e.g. a topological region) and physical entities (e.g. a table, a snowfall, etc.).

From the KCO module (http://www.loa-cnr.it/ontologies/KCO.owl) IRE imports:

$$\{ComputationalObject\}$$

$$ComputationalObject(x) \rightarrow InformationRealization(x)$$

The KCO ontology specializes the Information Objects ontology in order to build a conceptual schema for digital and analog content. The concept *ComputationalObject* specializes $InformationRealization$ for the computational world. Any physical document, electronic service, file, application, etc. are considered here computational objects.

4.2 The IRE predicates and axioms

Based on those predicates, IRE introduces the following predicates:

$$\{URI, AbstractWebLocation, ResolutionMethod, Resource, WebResource, ProxyResource, SemanticResource, hasIdentifier, webLocationOf, proxyFor, resolvableProxyFor, nonResolvableProxyFor, approximateProxyFor, exactProxyFor, informalExactProxyFor, formalExactProxyFor\}$$

The following axioms either characterize or define the above predicates:

$$hasIdentifier(x, y) \rightarrow Region(x) \wedge xsd : Datatype(y) \quad (1)$$

The axiom 1 introduces a relation between identifiers (specializing the class *Region* from DOLCE) and datatypes as encoded in XSD. In OWL, this is a so-called DatatypeProperty.

$$URI(x) \rightarrow xsd : Datatype(x) \quad (2)$$

A URI is characterized in 2 as an XSD datatype, since this is the current practice. The possible integration between DOLCE regions and datatypes consists in assuming a datatype structure as a metrics for DOLCE regions; for example, $xsd : Date$ can be assumed as a metrics for a subset of 'time intervals', which are regions in DOLCE.

$$AbstractWebLocation(x) =_{def} AbstractRegion(x) \wedge \exists y(URI(y) \wedge hasIdentifier(x, y) \wedge \neg\exists z(URI(z) \wedge y \neq z \wedge hasIdentifier(x, z)) \quad (3)$$

Abstract web locations are defined in 3 as DOLCE abstract regions that have exactly one URI as an identifier. Furthermore, in 4, we state that no URI can identify more than one abstract region. This ensures the isomorphism between URI-like encoding and the abstract web space generated by it (the abstract web space being the maximal sum of all abstract web locations).

$$AbstractWebLocation(x) \rightarrow \neg\exists(y, z) (URI(y) \wedge AbstractWebLocation(z) \wedge x \neq z \wedge hasIdentifier(x, y) \wedge hasIdentifier(z, y)) \quad (4)$$

$$webLocationOf(x,y,t) =_{def}$$
$$eAbstractLocationOf(x,y,t) \wedge$$
$$AbstractWebLocation(x) \wedge$$
$$ComputationalObject(y) \wedge TimeInterval(t) \quad (5)$$

5 defines a relation between an abstract web location and a computational object at a time interval. The definition specializes the relation $eAbstractLocationOf$ (imported from the spatial relations ontology) for abstract web locations and computational objects. $eAbstractLocationOf$ is a generic relation holding between regions and physical objects.
Notice that being a web location of a resource does not imply the successful resolution of the URI for the abstract web location into the resource (see 7 below), but only the assignment of an address to the resource.

$$ResolutionMethod(x) \rightarrow Method(x) \quad (6)$$
$$Resource(x) =_{def} ComputationalObject(x) \wedge$$
$$\exists m(ResolutionMethod(m) \wedge involves(m,x)) \quad (7)$$

7 defines resources as computational objects that can be web-located in an abstract web location. The cardinality 0..* of $webLocationOf$ for computational objects in general is not relevant here for the definition of the concept Resource (the relation has that cardinality by default). On the other hand, in operational terms, resources have been here restricted to those computational objects that are involved in a computational method that ensures the resolution of a URI, given certain circumstances (e.g. an abstract web location is an assigned web location of the resource, the server is switched on, the connection is active, etc).

$$WebResource(x) =_{def} Resource(x) \wedge \exists(y,t)$$
$$(webLocationOf(y,x,t) \wedge AbstractWebLocation(y))$$
$$\wedge TimeInterval(t)) \quad (8)$$

8 defines resources that have an assigned abstract web location. The actual resolution of a URI into the resource is still dependent on given circumstances (e.g. the server is switched on, the connection is active, etc).

$$proxyFor(x,y,t) =_{def} Resource(x) \wedge Entity(y) \wedge$$
$$TimeInterval(t) \wedge \exists z(InformationObject(z) \wedge$$
$$realizes(x,z,t) \wedge about(z,y,t)) \quad (9)$$

9 defines a relation between a resource x and any entity y the resource can be a proxy for, at a time interval. The definition says that for x to be a proxy for y, it must *realize* an information object that is *about* y.
Aboutness is dependent on how an agent is able to internally represent y based on the available information. Therefore, the operational grounding of this relation depends on the properties of y, and on the technology that can enable an agent to reach y from x, and to make operations on it.
In the following definitions, we introduce a typology of proxy relations, independently on the available technology.

$$ProxyResource(x) =_{def} WebResource(x) \wedge$$
$$\exists(y,t)(proxyFor(x,y,t) \wedge Entity(y)) \wedge$$
$$TimeInterval(t)) \quad (10)$$

In 10 we define proxy resources, which are web resources that are proxies for some entity at some time.

$$SemanticResource(x) =_{def} ProxyResource(x) \wedge$$
$$\exists(y,t,z,w)(proxyFor(x,y,t) \wedge$$
$$InformationObject(z) \wedge FormalLanguage(w) \wedge$$
$$realizes(x,z,t) \wedge about(z,y,t) \wedge orderedBy(z,w)) \quad (11)$$

The definition 11 introduces semantic resources, i.e. the proxy resources that realize information objects that are ordered (namely, encoded) by means of a formal language, e.g. OWL [20].

$$resolvableProxyFor(x,y,t) =_{def} proxyFor(x,y,t) \wedge$$
$$ProxyResource(x) \wedge WebResource(y) \quad (12)$$

The definition 12 says that only a proxy resource can be a resolvable proxy for an entity y, if y is a web resource on its turn. This is the typical case of `href` links on the Web.

$$nonResolvableProxyFor(x,y,t) =_{def}$$
$$proxyFor(x,y,t) \wedge \neg(WebResource(y)) \quad (13)$$

The definition 13 says that any kind of resource is a non-resolvable proxy for an entity y, if y is not a web resource. This case of proxy does not allow the resolution of a URI for the resource to y, since y has no abstract web location. Notice that only computational objects can be resolvable, therefore even if a non-computational entity, such as a table, a tennis match, or Peter Pan, would be technologically reachable, e.g. by means of sensors, bar codes, or simulators, their reachable part or annex would be anyway a computational object, with a different identity.

$$approximateProxyFor(x,y,t) =_{def}$$
$$proxyFor(x,y,t) \wedge \exists z(Entity(z) \wedge y \neq z \wedge$$
$$proxyFor(x,z,t)) \quad (14)$$

A resource is an approximate proxy for an entity (14) when it is a proxy for at least two entities, e.g. an html page can contain more than one link, or a document (text, image, etc.) can refer to several entities, either computational or not.

$$exactProxyFor(x,y,t) =_{def} proxyFor(x,y,t) \wedge$$
$$\neg\exists z(Entity(z) \wedge y \neq z \wedge proxyFor(x,z,t)) \quad (15)$$

A resource is an exact proxy for an entity (15) when it is a proxy for only one entity, e.g. a (URI-referenced) entry from an online dictionary, an image of a simple object, or a semantic schema that provides information to something.

$$informalExactProxyFor(x,y,t) =_{def}$$
$$exactProxyFor(x,y,t) \wedge \neg(SemanticResource(x)) \quad (16)$$

A resource is an informal exact proxy for an entity (16) when it is not a semantic resource, i.e. when it is not encoded in a formal language.
For example, the URI-referenced entry of an Italian legal glossary: http://fooleg.it/glossary.xml#Controparte is an informal exact proxy for the concept *Controparte* (meaning 'party') in the Italian Law.

$$formalExactProxyFor(x, y, t) =_{def}$$
$$exactProxyFor(x, y, t) \wedge SemanticResource(x) \quad (17)$$

A resource is a formal exact proxy for an entity (17) when it is a semantic resource, i.e. when it is encoded in a formal language.
For example, the URI-referenced entry of an Italian legal ontology: http://fooformleg.it/ontology.owl#Controparte is a formal exact proxy for the concept *Controparte* in the Italian Law.

5. CONCLUSION

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. There is something rather ambitious that the web is supposed to allow than just referencing web resources; that is, referencing things in general. On the other hand, that goal requires a software agent in order to identify a resource in an unambiguous way, and to perform the appropriate operations on it.
We have singled out some ambiguities that could prevent successful solutions to the web identity crisis. We have found a way to clarify and formalized them in a unique modelling framework. Finally, we have suggested some extensions to the model that can help classifying syntactic and operational solutions, and verifying their completeness and consistency.
Currently, URIs are used as the uniform mechanism for identifying heterogeneous entities, e.g., documents, metadata, ontologies, abstract concepts, physical things, events. But there is no clear categorization of which are the possible ways to identify and reference those entities on the web. We support the use of a formal model for the categorization of the things that can be referenced on the web. To this aim, we have defined IRE (Identifiers, Resources, and Entities), a conceptual pattern based on an ontology of Information Objects [1], built on top of DOLCE and on some of its modular extensions [15, 9, 8, 14]. IRE provides the framework to define a categorization of the kinds of resources that can be referenced on the web. We are confident that, based on this categorization, it is possible to study the most suitable way of handling the operational semantics that can be applied to different references. In particular, we have identified four kinds of relations that might hold between resources and entities (i.e., *proxyFor*) each depending on the degree of formality of the resource. This approach does not grant a comprehensive solution to the process of unambiguously identifying a resource, but puts a theoretical foundation to different suitable 'lower level' approaches. On the web there are different ways of identifying entities referenced by a resource. In our model this is reflected by the relations $resolvableProxyFor$, $approximateProxyFor$, $informalExactProxyFor$ and $formalExactProxyFor$. The first specifically refers to the retrieving of a computational object on the web (an hypertext link is a proxy for the document it allows to retrieve), the next two are informal way of identifying an entity, they rely on the support of a human agent (e.g., a personal home page). The latter provides support to the automation of the task (e.g., a semantic resource that is processable by a software agent which can infer facts on the described entity) and is compliant with the direction undertaken by the semantic web vision.

6. REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

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

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

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

[8] 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.

[9] A. Gangemi, N. Guarino, C. Masolo, and A. Oltramari. Sweetening wordnet with dolce. *AI Mag.*, 24(3):13–24, 2003.

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

[11] 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.

[12] G. L. M. and M. G. ISO/IEC 13250-2: Topic Maps – Data Model, 2005. http://www.isotopicmaps.org/sam/sam-model/.

[13] M. K. Smith, C. Welty, and D. L. McGuinnes. OWL Web Ontology Language Guide. Technical report, W3C, 2004. http://www.w3.org/TR/2004/REC-owl-guide-20040210/.

[14] 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.

[15] C. Masolo, L. Vieu, E. Bottazzi, C. Catenacci, R. Ferrario, A. Gangemi, and N. Guarino. Social roles and their descriptions. In D. Dubois, C. A. Welty, and M.-A. Williams, editors, *KR*, pages 267–277. AAAI Press, 2004.

[16] 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/.

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

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

[19] 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/.

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

[21] 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).