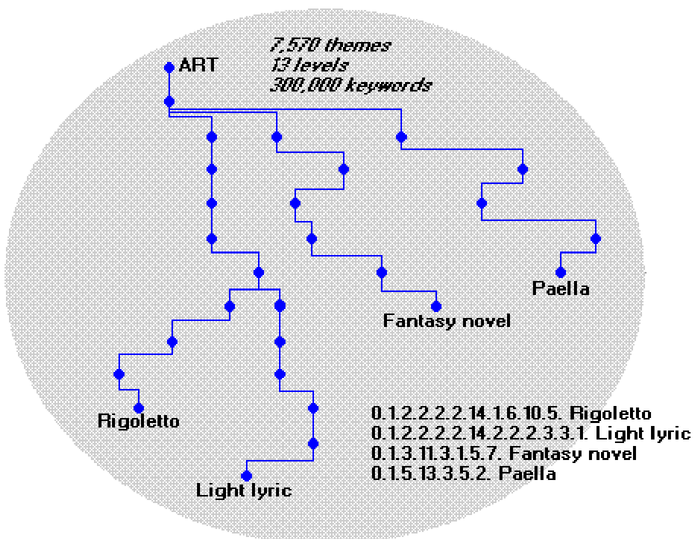
Web Semantic

Going from Mind to Digital or from Digital to Mind?

W3C versus Darwin, two polar approaches that could reinforce each other

Dr. Juan Chamero, jchamero@caece.edu.ar, Buenos Aires, Argentina February 15th 2009

The Darwin Vision



In the figure above is depicted a series of "**semantic paths**" highlighted in an **ART Map** unveiled from the **"Web as_it_is"** at a given moment by Darwin agents working within its particular **Darwin ontology**. This map extends from "root" to "leaves" throughout 13 "semantic levels". ***Rigoletto*** as name of the well known Italian opera is linked to the ART root by a **"semantic path"** of 11 levels [0.1.2.2.2.2.14.1.6.10.5] meaning the semantic hierarchic embedding:

ART => The Art => Performing Arts => main Performing Arts => Theater => genres => Opera => Opera History => Italian Opera => Bel Canto Movement => Rigoletto

The paths depicted belong to a prototype built on August 2008. A sample of it may be downloaded from **Intag**, the Darwin proprietary Website. Once a discipline like ART is mapped all Web documents somehow dealing with ART could be tentatively cataloged over this semantic skeleton. And we say "tentatively" because maps unveiled may evolve along the time differentially in both its content and its structure.

"Give to Caesar what is Caesar's, and to God what is God's"

The ambitious W3C project also known as the **World Wide Web Consortium** is now showing signs of weakness and tiredness. However we believe that its findings are all valid and with a promising future to cope efficiently with the future Web needs. It was conceived by Tim Berners Lee the Web father. The best universities, labs, and people worked hard since the beginning of the 90´s under the W3C big umbrella creating all type of abstracts tools and languages to implement TBL central idea: **The Web Semantics**, its personal and revolutionary vision, *an information space to be used by machines rather than by humans*. Instead of processing and manipulating Web information, users are encouraged to send agents to catch specific information, eventually forms of intelligence and to solve problems. To conceive its logical skeleton he uses the OSI Seven Layers Model, successfully implemented in the Internet beginnings. See Appendix below.

The imperfect human semantic paradox

Supposing that all information generated by humans or mechanically generated under their control is readable, fully "understood" and cataloged by machines all W3C languages and tools would work as a Swiss watch soon, let´s say in no more than a decade. The problem is that this information is actually huge, too much unstructured, fuzzy and noisy and perhaps in large extent these last apparent drawbacks are paradoxically necessary!. Concerning this paradox we advise to see our "Semantic Search" White Paper where we discuss about ideas, ideals and concepts, as a philosophical approach to build digital ontologies.

As we are going to analyze a little deeper below it seems that W3C project proceeded from digital to mind abandoning preliminary and necessary scouting tasks from mind to digital. Among these scouting tasks in the "Web Ocean" we may mention creating procedures to:

a) Detect and unveil keywords from existing documents;
b) Discover the main topic/s dealt with in documents:
c) Discover "authorities" and levels of authoritativeness;
d) Discover dominant logical semantic skeletons within the Web Ocean, namely hierarchically interrelated document clusters that share the same "main subject";
e) Discover the fine hierarchical structure of these apparently -at first sight- homogeneous clusters;
f) Discover invariant meanings within each language (concepts);
g) Discover for each knowledge domain its modal "namespace" (by the way you may find here a deep difference between W3C and Darwin approaches: objects namespace in W3C are fictitious but considered as true if referred to a given Website taken as namespace authority meanwhile Darwin agents try to get modal names statistically that is the "best" name for any meaning at a given time within the Web Ocean);
h) Unveil dominant "well structured documentation patterns", namely how top authorities document their ideas.

In brief we could say that our mind is too complex and that we are also in the beginning of the beginning of its understanding. We have problems to define what an idea is, what a concept, an ideogram or a single keyword. Another problem we have to take into account when dealing with networks of Internet type is that man-machine interactions are performed in two domains instead of one: the Web documents' side and the people's side. W3C approach ignores this fact. It works more with metadata that with data and supposing that data is always clean and "well structured". Darwin on the contrary works more with data than with metadata and it also works on the people´s side data trying to unveil the hidden semantic of real interaction queries.

In terms of the Theory of Computing Complexity we were performing some experimental computations using the "Blum´s Axiomatic Approach" arriving to some preliminary conclusions that could be expressed as follows: if the OSI Seven Layers Model had a complexity of 10 powered at 15 a similar model trying to approach to the human mind would have a complexity of 10 powered to 19, four more orders.

The second best

The pragmatic IT people who needs to solve hot problems applause and try to use those W3C tools and languages that consider useful now and there like **XLM**, **RDF** and some ideas of **OWL.** The applications people patch most of their semantic needs with programs and scripts written in C++, Java and .NET by "de facto" intertwining at their best wisdom thru the TBL seven abstract layers but not systematically but with punctual approaches far from universal, following by de facto a second best strategy. This circumstantial practical approach shall not invalidate W3C tasks, findings and ideals that should continue but with a different expectancy of observance

Wrongly used and instrumented, W3C findings, without being fully aware of its actual limitations could be costly and dangerous as long as government agencies, universities, professional associations, and corporations were induced to implement "sine qua non" their approaches. It will imply to give priority to the form instead of to the essence. On the contrary if used as trends and ideals to be met as much as possible it will be like the big lantern of the Web evolution.

W3C tools and languages used without common sense would resemble Ptolomeic versus Galileo Galilei approaches. Following W3C schemas to describe and solve actual applications would mean to fill thousand of description sheets and forms before start programming even trivial applications. And once everything is ready for programming the debugging task appears in the horizon as a real menace. What happens is that in scenarios of machines trying to understand, emulate and manage other machines, in order to coordinate common work among them, nothing is understood. For example a typical bureaucratic procedure in a government agency will render hundreds of steps thru hundreds of possible office “cubicles”. However human programmers advised and guided by red tape experts will go thru specific shortcuts of a few steps thru a small set of cubicles. It means programs with only a few control points to be checked versus hundreds, thousands for routine applications. That´s the cost of universality many times unnecessary and out of scale..

Darwin versus Tim Berners Lee vision

This concept was born in the “**The Semantic Web**” article published in the ***Scientific American*** on May 2001 by ***Tim Berners Lee***, the Web creator. For him the Semantic Web is an information space used by *machines* rather than *humans*. Instead of imagining processing and manipulating Web information via humans Lee imagined users sending their agents to the Web Ocean instead but proceeding systematically thru layers. This definition is oriented to a man – machine Communications realm but necessarily it should be coherent with a human–human communication system and embedded in its turn as a macro layer in an augmented human–machine–human communication system. Under this augmented and more realistic point of view the Web Semantic could be considered the study of “meanings” of all semantic creatures dispersed over the **Web Ocean**: documents. Darwin “sees” the **SW** as a binary coupled interacting system of two subsystems: the Web Ocean and Humans trying to enhance their knowledge in order to survive and evolve properly, in single words a **huge collective learning system**. Some humans to be known as “authors” put their seeds (documents) now and there in the Web Ocean motivated either by a natural passion to transcend or by solidarity or by both. These creatures are not alive like fishes are; they resemble “fossils” of a piece of knowledge because their meanings are frozen at the time of their sowing. However these pieces of knowledge have a “cognitive life” in the sense of their usefulness for other human beings. So from the point of view of cognition they could be considered equivalent lo living creatures except that they do not move by themselves. For this reason Darwin ontology states that the whole stock of these Web creatures could be considered a sample of the **Human Knowledge** as frozen at a given time and named a sample of the “**Established Knowledge**” at a given time as well. So the **EK**, Established Knowledge should be considered a fossil in the sense that its validity is from the past to the present and only as a trustable Knowledge basement for the future. Where is then the seed of the future?. Where are then the new ideas and the necessary updating of everything?. **Darwin ontology** states that it rests on the people realm as Web users. In ***Kant*** and ***Spinoza*** terms the Web Ocean is a fiction of the Human Knowledge and the closest to their vision of truth rests on “people´s side”, users that interact thru Internet protocols with the Web Ocean. In this side ideas and potential seeds are continuously generated and bubble within people´s minds. People have at their reach a huge, universal, open and free learning system: the Web Ocean where the best “modal ideas” and meanings at a given time rest.

TBL Vision and W3C in more detail

This Consortium also known as World Wide Web or W3C is devoted to the creation of languages, tools and standards to implement the WS properly. Let´s make an introduction to them.

XML: Extensive Markup Language, a syntax model to define the content structure of documents without taking care/worrying about their meaning!. Derived from this syntax and even using XML specialized editors may be created such as **XHTML**, **MathML**, and **MusicML**, oriented to the Web, Mathematics and music respectively. The following XML file is used in various samples throughout the Microsoft XML Core Services (**MSXML**) **SDK.**

```
<?xml version="1.0"?>
<catalog>
  <book id="bk101">
    <author>Gambardella, Matthew</author>
    <title>XML Developer's Guide</title>
    <genre>Computer</genre>
```

```
    <price>44.95</price>
    <publish_date>2000-10-01</publish_date>
    <description>An in-depth look at creating applications
    with XML.</description>
  </book>
  <book id="bk102">
    <author>Ralls, Kim</author>
    <title>Midnight Rain</title>
    <genre>Fantasy</genre>
    <price>5.95</price>
    <publish_date>2000-12-16</publish_date>
    <description>A former architect battles corporate zombies,
    an evil sorceress, and her own childhood to become queen
    of the world.</description>
  </book>
………………………………………………….
……………………………………………..
</catalog>
```

It seems us that no further explanation is necessary for programmers to understand the XML nature.

XML Schema: is a language of schemes to define precisely all type of XML documents. It states how an XML documents should be. We depict below an example taken from W3CShcools to define precisely a shipping order:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<shiporder orderid="889923"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="shiporder.xsd">
<orderperson>John Smith</orderperson>
<shipto>
<name>Ola Nordmann</name>
<address>Langgt 23</address>
<city>4000 Stavanger</city>
<country>Norway</country>
</shipto>
<item>
<title>Empire Burlesque</title>
<note>Special Edition</note>
<quantity>1</quantity>
<price>10.90</price>
</item>
<item>
<title>Hide your heart</title>
<quantity>1</quantity>
<price>9.90</price>
</item>
</shiporder>
```

The XML document above as a logical sub-tree consists of the root element, "shiporder" that contains a required attribute called "orderid". It is meaningful for any employee that has issued a shiporder. The "shiporder" element contains three different derived elements: "orderperson", "shipto" and "item". The "item" element appears twice, and it contains a "title", an optional "note" element, a "quantity", and a "price" element.

Note: The line above: xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" tells the XML parser that this document should be validated against a schema. The line: xsi:noNamespaceSchemaLocation="shiporder.xsd" specifies WHERE the schema resides.

Even though this is a Little more complex it would be easily understood by any smart programmer specially if aided/guided by an XLMSchema constructor/editor. Up to herein I´m prone to applaud these two W3C initiatives as an acceptable quote of necessary IT&C “red tape”. .

RDF: Resource Description Framework is a language to model interrelated data at their turn presented in XML syntax. Let´s see the RDF example of the same school applied to a list of CD´s for sale:

Title	Artist	Country	Company	Price	Year
Empire Burlesque	Bob Dylan	USA	Columbia	10.90	1985
Hide your heart	Bonnie Tyler	UK	CBS Records	9.90	1988

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:cd="http://www.recshop.fake/cd#">
<rdf:Description
 rdf:about="http://www.recshop.fake/cd/Empire Burlesque">
  <cd:artist>Bob Dylan</cd:artist>
  <cd:country>USA</cd:country>
  <cd:company>Columbia</cd:company>
  <cd:price>10.90</cd:price>
  <cd:year>1985</cd:year>
</rdf:Description>
<rdf:Description
 rdf:about="http://www.recshop.fake/cd/Hide your heart">
  <cd:artist>Bonnie Tyler</cd:artist>
  <cd:country>UK</cd:country>
  <cd:company>CBS Records</cd:company>
  <cd:price>9.90</cd:price>
  <cd:year>1988</cd:year>
</rdf:Description>
.
.
.
</rdf:RDF>
```

With an almost straightforward “natural” meaning:
First line => It´s an XML declaration
Second line => the root name of declaration RDF
As there were rdf names with specific names as in any programming language lines 3 and 4 tell us that the RDF names must belong to the xmlns (XML name space) located in a specific file of the W3C portal and those of the particular space to which this declaration belongs will be located in a “fake” Website that supposedly deals with CD lists.
Line 5 uses the first name of the rdf space: “description” and we are aware of this fact because the prefix structure above mentioned rdf:Description about (a second permitted name) with an argument that tells us where is the CD theme located and its name “Empire Burlesque”.
Lines 6 to 10 are now easy to imagine: CD description by “artist”, “country”, “company”, “price”, and “year”.
Once finished a CD description follows next in the same way.

The same conclusions about the utility of XML and XMLSchema are valid for this tool. At large any smart programmer does pretty much the same and for this reason to encourage them to invest a little more of their time to use these tools whether they become universal and sufficiently tested will bring benefit to all the IT&C community. However it (becoming universal and sufficiently tested) will take its time.

Note: to see the status of this core W3C package (the necessary and sufficient trilogy XML, RDF, OWL?) we recommend to read The Scientific Publishing Task Force report

OWL: Ontology Web Languages, is a family of languages to represent the knowledge even tough restricted to the domain of digital computation and communications, A Web ontology is a set of axioms and/or conjectures that are valid fort certain “classes” of the Web space.

OWL is designed for use by applications that need to process instead of structured objects for which meanings are known. It´s supported by previously defined above tools and at its turn it opens in three increasingly-expressive sublanguages: **OWL Lite**, **OWL DL**, and **OWL full**.

It´s important to emphasize here again that also OWL is essentially descriptive. For instance OWL Life provides efficient migration for" Thesauri and other taxonomies". So it departs of knowing the Web Thesauri which has not being extracted from the Web yet!. We are facing the classic paradox of the egg and chicken: who is first. Concerning the Web we affirm undoubtedly the egg namely the Web Thesaurus.

OWL DL deals with all possible forms of computable logic in order to approach as much as possible to human thinking. Finally **OWL full** tries to cover all types of human text corpuses but without assuring their computability.

Darwin versus OWL approaches

An early (about "wine") OWL example will illustrate us better that continuing talking in abstract. This example was built to guide the Web search of an special agent, the KSL Wine Agent from Knowledge System Laboratory, Computer Science Department, University of Stanford. .

That´s nice because at the same time it give us the chance to compare both approaches about to know and manage the Web Semantics: The creation of a complete logical mathematic semantically oriented infrastructure (W3C) or accepting the Web ocean as_it_is today and "sine die" , a huge reservoir of Web documents written at will without trustable descriptors, in any language and in any form.

The first effort is necessary and it will be especially useful to discipline the way we humans document and to build precise baits and agents to retrieve specific pieces of knowledge as it was experimentally tested with the agent above commented. The other way is trying to create ontologies that interpret at least statistically and reasonable well the way we humans document ideas. Darwin ontology was created in the belief that this task is mentally performed thinking in the neighborhood of "modal"/dominant ideas. In my humble opinion both ways are valid and at large complementary.

Let´s suppose that all our efforts go along the first approach without trying first to detect and unveil meanings and then to proceed to "order" them by "meaning". Before going on blindly we should have to ask ourselves: how many meanings could be dispersed on the Web Ocean?. The answer probably rounds billions. W3C tools enable us to build agents and baits to match and catch efficiently each meaning as long as documents dealing with them were written **as_it_should_be** adjusting their form to W3C standards as much as possible. On the contrary, following the second approach we may detect and unveil all possible meanings and classify them hierarchically, once and forever from an evolutionary point of view. This necessary primal ordering will facilitate and make worthy the W3C effort and/or similar.

Wine OWL Example

According to the specification, a wine is a potable liquid produced by at least one maker of type winery, and is made from at least one type of grape (such grapes are restricted to wine grapes elsewhere in the ontology.). The full text of the declaration additionally stipulates that a wine comes from a region that is wine-producing and, most importantly to the agent, that a wine has four properties: color, sugar, body, and flavor.

If you are a programmer you may easily guess what the precise OWL Wine example text corpus tell us and for your control check block by block with the W3C description below:

*"The ontology defines the concept of a wine with the declaration excerpted on the right. According to the specification, a wine is a **potable liquid** produced by **at least one maker** of **type winery**, and is made from at **least one** type of **grape** (such grapes are restricted to **wine grapes** elsewhere in the ontology.). The full text of the declaration additionally stipulates that a wine **comes from a region** that is **wine-producing** and, most importantly to the agent, that a wine has **four properties**: **color, sugar, body**, and **flavor**.*

Note: It appears here a sort of alien command "daml" who stands for DARPA Agent Markup Language being DARPA a military US organization, ***Defense Advanced Research Projects Agency***, one of the W3C founders and main sponsor.

DARPA minCardinal is a sort of neologism that talks about existence in the logical use of our common expression “at least”. It is really a bit more complex than that we don´t err too much by using this equivalence.

```
<rdfs:Class rdf:ID="WINE">
    <rdfs:subClassOf rdf:resource="#POTABLE-LIQUID"/>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#MAKER"/>
        <daml:minCardinality>
          1
        </daml:minCardinality>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#MAKER"/>
        <daml:toClass rdf:resource="#WINERY"/>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#GRAPE-SLOT"/>
        <daml:minCardinality>
          1
        </daml:minCardinality>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#GRAPE-SLOT"/>
        <daml:toClass rdf:resource="#WINE-GRAPE"/>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#REGION"/>
        <daml:minCardinality>
          1
        </daml:minCardinality>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#REGION"/>
        <daml:toClass rdf:resource="#WINE-REGION"/>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#SUGAR"/>
        <daml:minCardinality>
          1
        </daml:minCardinality>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#SUGAR"/>
        <daml:toClass rdf:resource="#WINE-SUGAR"/>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#FLAVOR"/>
        <daml:minCardinality>
          1
```

```
        </daml:minCardinality>
      </daml:Restriction>
```

```
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#FLAVOR"/>
        <daml:toClass rdf:resource="#WINE-FLAVOR"/>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#BODY"/>
        <daml:minCardinality>
          1
        </daml:minCardinality>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#BODY"/>
        <daml:toClass rdf:resource="#WINE-BODY"/>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#COLOR"/>
        <daml:minCardinality>
          1
        </daml:minCardinality>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#COLOR"/>
        <daml:toClass rdf:resource="#WINE-COLOR"/>
      </daml:Restriction>
    </rdfs:subClassOf>
  </rdfs:Class>
```

Complementary W3C tools and languages

SPARQL: an RDF Query language. Sparql is a **recursive acronym** that stands for Protocol and RDF Query Language. Let's see an example:

```
PREFIX abc: <http://example.com/exampleOntology#>
SELECT ?capital ?country
WHERE {
  ?x abc:cityname ?capital ;
     abc:isCapitalOf ?y .
  ?y abc:countryname ?country ;
     abc:isInContinent abc:Africa .
}
```

It is a query that renders us all country capitals in Africa.

This language uses **"triples"** under the form of **subject – predicate – object**: Tony is tall; John loves Maria. Commonly used triples count by billions. The application of this advanced tool is not trivial. It needs of: a) servers specially adapted to RDF and SPARQL; b) common metadata and maps based on particular ontologies (like Darwin Maps); c) The WS set of rules, known as RIF, Rule Interchange Format; d) hyperlinks and "Meta tags" enabling the automatic semantic hyperlinks generation: special family of agents to perform the different tasks; and e) Web Services to be operated by agents.

W3C Application Projects under way

Neurocommons: they have an RDF database extracted from important Science of Life databases focused on neuroscience and using SPARQL. They say about themselves:

The NeuroCommons project seeks to make all scientific research materials - research articles, annotations, data, physical materials - as available and as useable as they can be. We do this by both fostering practices that render information in a form that promotes uniform access by computational agents - sometimes called "interoperability". We want knowledge sources to combine meaningfully, enabling semantically precise queries that span multiple information sources.

FOAF, Friend of A Friend: a machine readable ontology describing persons. See an example of a personal profile:

```
<rdf:RDF xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <foaf:Person rdf:about="#JW">
    <foaf:name>Jimmy Wales</foaf:name>
    <foaf:mbox rdf:resource="mailto:jwales@bomis.com" />
    <foaf:homepage rdf:resource="http://www.jimmywales.com/" />
    <foaf:nick>Jimbo</foaf:nick>
    <foaf:depiction rdf:resource="http://www.jimmywales.com/aus_img_small.jpg" />
    <foaf:interest rdf:resource="http://www.wikimedia.org"
          rdfs:label="Wikipedia" />
    <foaf:knows>
      <foaf:Person>
        <foaf:name>Angela Beesley</foaf:name>
        <!-- Wikimedia Board of Trustees -->
      </foaf:Person>
    </foaf:knows>
  </foaf:Person>
</rdf:RDF>
```

SIOC, Semantically Interlinked Online Communities. SIOC provides methods for people to people interaction such as blogs, forums and mailing lists. Its vocabulary is based on RDF and is defined using RDFS.

SIMILE, it stands for Semantic Interoperability of Metadata and Information in unLike Environments, Massachusetts Institute of Technology. As per its authors "*It is focused on developing robust, open source tools that empower users to access, manage, visualize and reuse digital assets ".* It has been applied to many uses, for example: **Appalachian**: a Firefox add-on that adds the ability to manage and use several OpenIDs to ease the login parts of your browsing experience. **Timeline**: A DHTML AJAX timeline widget for visualizing temporal information. Take a look here to open your mind about the wide spectrum of amazing W3C applications!.

Linking Open Data Community Project: it is a community effort to create open RDF databases to be interconnected. Among many others are: **DBpedia** containing 2,180,000 concepts extracted from Wikipedia and describe 218,000,000 triples in 11 languages; **DBLP Bibliography** containing 800,000 articles, 400.000 authors and 15,000,000 triples; **GeoNames**, providing RDF geographical description of more than 6,500,000 locations and geo-accidents worldwide; **nese**, the first European people statistic information with 500,000 entries deployed with XHTML and RDF.

Some other related links

- http://www.semanticweb.org/wiki/ISWC2008
- http://iswc2008.semanticweb.org/
- http://www.appliedminds.co.uk/

Towards a Synthesis

"Give to Caesar what is Caesar's, and to God what is God's"

The Semantic Web utopia is a formidable Project. However W3C languages and tools created to make it possible are not yet tested and far from being universal, friendly and efficient. As we have said in the introduction of these reflections W3C does not care about how humans document their ideas but documents are the unique cognitive creatures of the Web Ocean as_it_is today. With unlimited time the W3C tools and languages will work as planned because they ideally enable and facilitate humans to document meanings properly resembling perfect WFF's Well Formed Formulae of the Logic and also to "catch" via agents arbitrary pieces of knowledge out of the Web Ocean as long as they were documented as WFF's. This strategy is advisable for scientific and technical documents issuing from now on and to universally manage scientific and technical databases as well but what to say about the actual knowledge asset dispersed as_it_is today in the Web ocean?. This knowledge is perfectly readable and understandable by humans even though not easily retrieved out of the Web Ocean because as of today documents are not semantically indexed.

Agents should fit humans not the contrary

In order to find a "middle way" strategy to solve this "ab initio" inconsistency we may adopt the following hypothesis: humans document their ideas (ab initio also) reasonable well (following some rules and criteria) statistically. It means that we may define for any discipline of the Human Knowledge a "core" of **WWD's, Well Written Documents** (adjusted to a given set of rules and criteria). *The rest within each discipline "surrounds" semantically this core, adding ambiguity, fuzziness, noise, an why not?: not yet established meanings and potential seeds of wisdom. The first need is then to create agents smart enough to locate these cores and to understand them. Agents should fit humans not the contrary. This is a little more that common sense!.*

Ah the imperfections and differences!

Imperfections ad differences are the spice of life. It´s like the beauty, a circumstantial and "modal idea" surrounded by a probabilistic weighted neighborhood of acceptable forms of beauty ideally displayed onto a continuum. We have to remember that we are evolving creatures and as such we are subject to the evolution wheel. Evolution needs of differences in order to create new things and specially concepts when restricting ourselves to the cognitive realm. We also believe that the Web as_it_is today is a treasure that has to be preserved and that its nurturing via modal ideas and differences encouraged as much as possible.

The beginning of the beginning

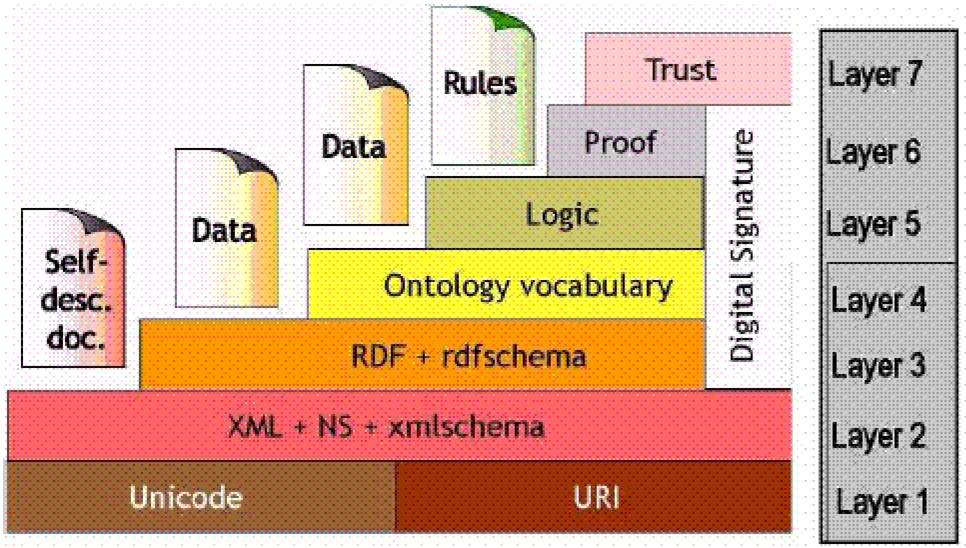
W3C was founded in October 1994, not too much in terms of organization life but old enough in terms of Web Development. W3C worked along almost 15 years going from "digital to mind" but no significant efforts were performed going from, "mind to digital". This asymmetry is dangerous and costly for the IT&C community. Along the need to reestablish the equilibrium the beginning of the beginning would be to build the **Web Thesaurus**, paradoxically the first proposed Tim Berners Lee utopia. It means to unveil semantically the Web as_it_is at a given moment: for example today. It means to hierarchically map all existing meanings per language. In numbers the recognition of more than 170 main Human Knowledge domains, totaling from 10 to 15 million meanings per language hierarchically distributed along 300,000 to 400,000 subjects.

The meaning of meanings – next steps

This is an intentional paradox. Meanings as objects must be understood by our mind which has the ability of recognizing meanings by using the verb "mean": for example I mean "that". So why not to imagine meanings as complex symbols of knowledge, let´s say something like symbols belonging to a "code". If in Theory of Information when Bob transmits Alice [1011], Alice agrees she has received four bits of information and as a result of it she is acquainted that Bob wanted to communicate her that the "answer" for a given problem or information she had is somehow and somewhere in the box number 11 of her internal or external memory why not to extend the analogy?. When a person say I mean "that" why not to imagine that he/she has received a flash of information that brings of a sudden to his/her memory a vision of a specific meaning, perhaps one out of millions?. We all manage reasonably well information, knowledge and intelligence as well as ideas and concepts..Our intuition suggests us that knowledge is a sort of fluid subtler than information and different from intelligence. When we say that an agent unveils information and knowledge we intuitively "know" what we mean by that difference. Information is related to pointing to something, many times the precise location of a piece of knowledge meanwhile knowledge is related to substance, to a necessary nutrient of our mind. When we say that an agent unveiled the hidden intelligence we mean that we have envisaged something like an structure, a frame, perhaps containing and specifying forms and modes of pieces of knowledge.

Appendixes

I- Tim Berners Lee Seven Layers Model



May 17th 2001 Tim Berners Leee Seven Layers Model: The Semantic Web, Scientific American

Sources:
Introduction to the Semantic Web Vision and Technologies, Part 2, Foundations, from Cody Burleson;, as Oct 4th 2007
A Semantic Web Status Model, by Aurona Gerber, Meraka Institute, from South Africa,

Layer 1

Layer 1 comprises Unicode and URI (Uniform Resource Identifier) technologies. The function of these technologies is to provide a unique identification mechanism for upper language technologies.
Unicode: Unicode aims to uniquely identify the characters in all the written languages by assigning a unique number to each character. The Unicode Standard (*Universal Character Set, Unicode/ISO10646*) specified by the Unicode Consortium is the universal character encoding standard used for representation of text for computer processing. This standard supports three encoding mechanisms, UTF-8, UTF-16 and UTF-32, allowing the same data to be encoded in a byte, word or double word format. The emergence of the Unicode standard and the availability of supporting tools are amongst the most significant recent global software technology trends. Unicode replaces the use of legacy character sets and it allows data and text to be exchanged internationally between different systems.
URI: A URI (Uniform Resource Identifier), defined as a compact string of characters that can be extended, is used to identify an abstract or physical resource. A *resource* is defined as an entity

that has identity. The general URI specification of the IETF (Internet Engineering Task Force) is known as RFC2396. URLs (or Uniform Resource Locators) are a subset of URI that specifically identify resources by using their network 'location' rather than identifying the resource by name or by other attributes. The Semantic Web would be impossible without global identification and hence the use of URIs. The future expansion of URIs into IRIs will ensure that a resource can be identified across language and character encoding boundaries and any discussion about 'meaning' has to uniquely identify the objects or resources of the discussion.

Layer 2

Layer 2 comprises of Namespaces, **XML** (Extensible Markup Language) and **XML Schema** technologies (Figure 1). The function of these technologies is to provide a self-describing syntax for the upper layer language technologies. We acknowledge the existence of DTD (Document Type Declaration) as XML originally used DTD as a validation mechanism, however, DTD was replaced by XML Schema and a discussion of DTD is thus excluded.

Namespaces: Namespaces (NS) provide a simple method for qualifying element and attribute names used in XML (Extensible Markup Language) documents. Namespaces are identified by URI references. The W3C Namespace Recommendation [16] defines an XML namespace as a collection of names, identified by a URI reference [RFC2396], which are used in XML documents as element types and attribute names.

XML: XML (Extensible Markup Language) specifies a standard for the exchange of data over networks, notably the Web. XML is considered to be both a *metalanguage* and a *markup language*. XML as metalanguage allows for the specification of the content of documents according to a predefined and specific structure. All documents conforming to this specification will have the same structure or represent data items in the specified structure. In addition, XML as markup language allows for the insertion of markup tags into text to define the logical structure of a document, or to add information regarding information contained in a document (metadata).

An XML document is a *text* document which in itself does not have any functionality. It is used only to describe data, information or metadata. Thus, XML is a mean for defining common grammars to enable data exchange. It does not specify semantics; all parties *must agree* on the data model and document structure for XML data exchange to be successful. If an XML grammar is accepted as a standard for data exchange, any XML parser can parse the XML data and access the content if it is a valid XML document. It is however difficult to re-engineer the data model from any given XML document if the document type specification is not available.

XML Schema: An *XML schema* is an XML document defining the content and structure of one or more derived XML documents. Generally, a *schema* is a model for describing the structure and content of data. XML Schema is a content modeling language as well as an application of *XML* that applies only to XML-related languages and documents. In particular, an XML Schema describes a model for a whole class of XML documents. The model describes the possible arrangement of elements, their attributes and text that would be present in a schema-valid document.

Layer 3

RDF (Resource Descriptive Framework) and RDF Schema technologies reside on Layer 3 (refer to Figure 1). The function of these technologies is to provide a metadata description mechanism for the upper language technologies.

RDF: The purpose of RDF (Resource Description Framework) is to declare metadata that is machine-readable,. RDF provides a mechanism to declare statements that describe resources by means of a basic data model. A *statement* describes an entity (resource) in terms of *properties*, which have *values*. Furthermore, an RDF statement is a [subject – predicate - object]. The *subject* is the resource of the statement. The *predicate* is the property or characteristic of the subject specified by the statement (examples include creator, creation-date, or language), and the value of the property is the *object*. In terms of the Semantic Web, the basic object-attribute-value data model is the only semantics prescribed in the RDF specifications. RDF has no other data-modeling A Semantic Web Status Model 5 commitments specifies no reserved terms for further data modeling or no other mechanisms for declaring property names. For semantic interoperability RDF has significant advantages over XML primarily because of the data model used.

RDF Schema: RDF Schema specifies extensions to RDF that are used to define the common vocabularies in RDF metadata statements. RDF itself provides the data model and does not prescribe any application-specific classes and properties. This is accomplished by RDF Schema. RDF Schema provides a predefined, basic type system for RDF models, thus extending RDF by assigning an externally specified semantics to specific resources. RDF Schema expressions are valid RDF expressions, and therefore RDF Schema is a semantic extension of RDF Software that can interpret RDF can also be used to interpret an RDF Schema implementation; although it will not attach the intended meaning to the built-in schema definitions.
The RDF vocabulary description strategy contained in RDF Schema acknowledges that there are many techniques that enable *description of meaning* of classes and properties. To extend the description of meaning, ontology languages (such as DAML+OIL, OIL and OWL), inference rule languages and other formalisms are used.

Layer 4

In Figure 1 'Ontology vocabulary' is depicted on Layer 4. Here the terminology differs from the three preceding layers, because the *functionality* rather than the *technology* is mentioned. OWL is the W3C technology representing this layer. **Ontology Vocabulary:** specifies a machine readable vocabulary in computer systems technology descriptions. Ontologies assist in creating a common understanding for communication between people and computer applications. Generally it is defined as a shared, formal, explicit specification or conceptualization of a particular domain. Furthermore, ontology is a *knowledge representation language* capturing the syntax (ontology vocabulary) as well as the semantics (logic language) of a specific domain,
It is envisioned that ontologies will play a crucial role in knowledge processing, sharing, and reuse between Web applications. On the Semantic Web, ontologies may be used in applications required to search across, or merge information from diverse communities.
In early 2001 the W3C initiated a Web-Ontology Working Group (WebOnt) in order to consolidate existing Web ontology efforts (notably **OIL** and **DAML+OIL**) into a Web Ontology Language. **OWL** extends RDF Schema in order to express complex relationships between different classes specified in RDF Schema, as well as to enhance the specification of constraints applicable to classes and properties. OWL specifies three sub-languages of increasing expressiveness. These language are **OWL Lite**, **OWL DL** and **OWL Full**. Ontology designers should select the most appropriate version.

Layer 5

The Semantic Web model depicts the 'Logic' layer as being *above* the 'Ontology vocabulary' layer. Logic or semantic descriptions are generally included in the specification of a knowledge representation language. Such languages are required to capture ontologies, as ontology typically specifies a domain concept hierarchy. A *knowledge representation language* is specified when both the syntax and the semantics of the language is described. In the syntax definition, the legal statements in the language are defined (using an ontology vocabulary), and the semantic description specifies the intended meaning of each legal statement.
Logic: The OWL language provides a specific subset in the form of **OWL DL** to support existing DL (Description Logics) and to provide a language subset that possesses the computational properties required for reasoning systems. DL is a set of knowledge representation formalisms with semantic characterization based on standard first-order logics. DL offers a formal foundation for frame-based systems, where meaning is provided by interpretations that define the formal semantics of the logic

Layer 6

Proof: 'Proof' as concept exists within the theorem proving domain, for instance as applied in artificial intelligence.. To support Semantic Web proof scenarios, *proof languages* were developed. A proof language determines the validity of a specific statement. An instance thereof generally

consists of a list of inference items used to derive the information in question, as well as the associated trust information of each item.
A Semantic Web will probably not require proof generation and in general proof validation will be adequate. The search for and generating a proof for an arbitrary question, is typically an intractable process for many real world problems, and the Semantic Web does not require this to be solved. For perceived Semantic Web applications construction of a proof is performed according to constrained rules, and only the validation thereof is required from other parties. For example, when a user is granted access to a Web site, an accompanying document explains to the web server why they should be granted access. Such proof for example, could be a chain of assertions and reasoning rules with pointers to all supporting material.

Layer 7

Trust: Semantic Web interaction requires different collaborators to communicate, implying that they have to determine how to trust one another, as well as how to establish the trust levels of acquired information. When dealing with user interactions on the Web, ***McKnight et. Al*** defined the term *trust* as the belief that another entity is benevolent, competent, honest, or predictable in a given situation. Trust also includes the participants' willingness to depend on one another in a specific interaction. Furthermore, user trust of Semantic Web information is determined by the *source* of the information, in particular its authenticity and trustworthiness.
Within the Semantic Web the concepts trust and proof are dependent on the interaction context. However, an all-encompassing definition of context is problematic. An appropriate meaning of context is therefore explicated by means of the following example: A user on the Semantic Web receives data from a friend regarding the best music performances. The data can be trusted as it originates from a *known* (implying verified) friend, whose musical interests are familiar. It is thus possible to *use* the data because the user either shares or disagrees with the musical tastes of the friend. Within the domain of the Semantic Web, *context* therefore assists applications or users regarding the trustworthiness and usefulness of data

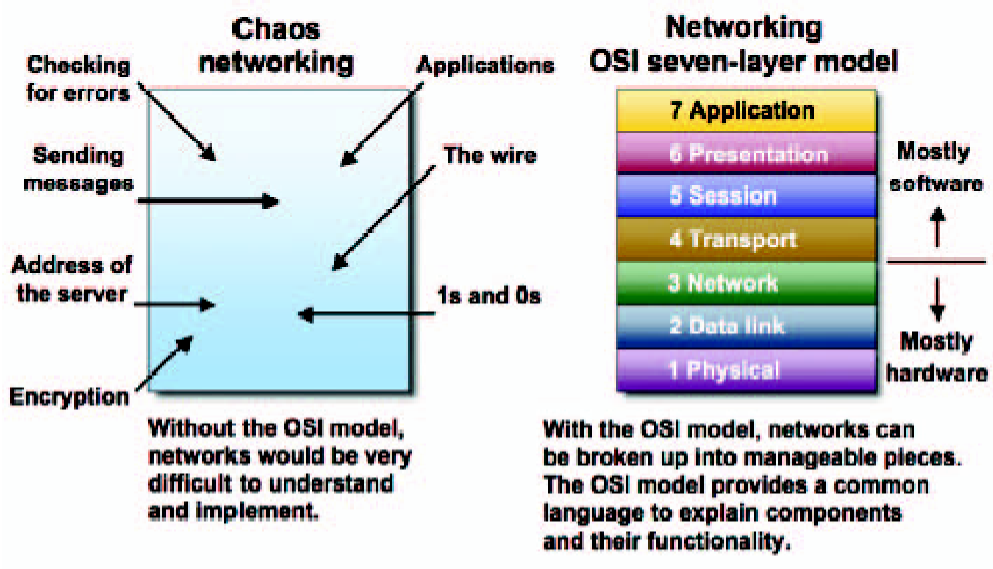
Digital Signatures

In the Semantic Web model of Berners-Lee **'Digital Signature**' is associated with the middle four layers. The **Digital Signature Standard** (**DSS)** is a cryptographic standard or a particular application of public key cryptography promulgated by ***NIST (National Institute of Standards and Technology)***

A digital signature (not to be confused with a **digital certificate**) is an electronic signature that can be used to authenticate identity. Digital signatures are easily transportable, cannot be imitated by someone else, and can be automatically time-stamped. A digital signature can be used with any kind of message, whether it is encrypted or not. A digital certificate contains the digital signature of the certificate issuing authority so that anyone can verify that the certificate is real.
On the Semantic Web a digital signature is a mechanism used to unambiguously verify an identity such as the author of a document. XML signatures are digital signatures designed for use in XML transactions. The implementation of digital signatures on the Semantic Web could result in a system which can express and reason about relationships across the whole range of public-key based security and trust systems.

II- The Seven Layers OSI Model

Sources: OSI Model, from Wikipedia; The 7 Layers of OSI Model, from Webopedia;



Source: GlobalKnowing.com, from its OSI Model White Paper

OSI Reference Model stands for **Open Systems Interconnection Reference Model,** an abstract description for layered man-machine communication. In its basic form it splits its architecture into seven layers, from top to bottom: the Application, Presentation, Session, Transport, Network, Data-Link, and Physical Layers.

A layer is a collection of conceptually similar functions that provide services to the layer above it and receives service from the layer below it. For example, a layer that provides error-free communications across a network provides the path needed by applications above it, while it calls the next lower layer to send and receive packets that make up the contents of the path.

7. **Application Layer**, the first (sending) or last(receiving) layer of a man-machine communication, in fact the closest to the end user. Examples: Telnet, FTP downloads and uploads; The POP3 process for email reception: a SMTP, Simple Mail Transfer Protocol; A Web page as “seen” by the HTTP protocol;

6. **Presentation Layer**, where code conversions are performed. It is also known as the Syntax Layer. Examples: EBCDIC to ASCII; Data marking and serializing objects, and/or encapsulations tasks previously to packet sending;

5. **Session Layer**, where connectivity and machines dialog is established. It is the layer responsible of the “graceful closing” of a session. Examples: full duplex, half duplex, or simplex; negotiations control of operations such as: updating; adjustments, restart procedures, repetitions, terminations, asking permissions, etc.

4. **Transportation Layer**, providing the necessary data transparency between users, segmentation control, errors handling and control of missing and rest of messages transmission. Examples: TCP Transfer Control Protocol and UDO, User Datagram Protocol. Their tasks are comparable to the ones of a conventional Post Office.

3. **Network Layer**, executes under its control variable length data sequences transmissions towards their destinies thru one or more networks. It also controls the whole quality of this service. It is the layer where routing, fragmentation and data assembly tasks are performed reporting errors whether existent. The best example of this complex function is the IP, Internet Protocol where data packages are fragmented as a function of the accessibility and bandwidths at hand.

2. **Data Link Layer** provides all means to internetworking data transmission, detecting and whether possible correcting errors generated in or by the inferior layer. Examples: A point to point connection between LAN's or WAN's.

1. **Physical Layer**, where all physical and electrical specification of all devices participating in the transmission, are matched and coordinated namely pins, voltages, wire and wiring specifications, hubs, network adapters, buses adapters, etc. It takes care of analog to digital and reverse conversions, modulation and demodulation, digital representations, analog converters, etc.