

## The Semantic Web

Alun Preece

Computing Science, University of Aberdeen  
(from autumn 2007: School of Computer Science, Cardiff University)

- ◆ What is the Semantic Web, and why do we need it now?
- ◆ How does the Semantic Web relate to the "traditional" Web?
- ◆ What are the main components of the Semantic Web architecture?
- ◆ What does the Semantic Web technology buy us in terms of applications



## What is the Semantic Web?



"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

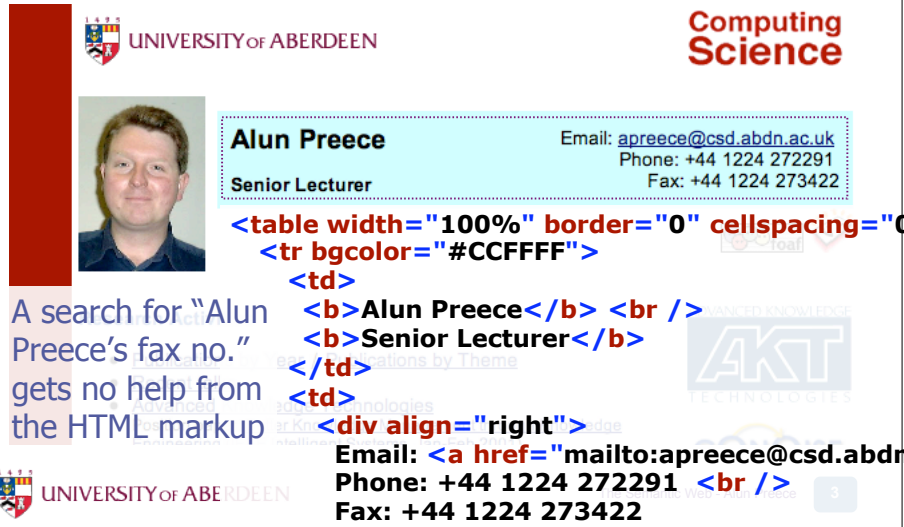
Tim Berners-Lee, James Hendler, Ora Lassila,  
*The Semantic Web*, Scientific American, May 2001

- ◆ Aim: to create a network of **machine-processable resources**
- ◆ Existing in parallel with the current World Wide Web
- ◆ Enables software to carry out **tasks** on users' behalf
- ◆ Moving from a Web of "finding things" to a Web of "doing things"



## The problem with the HTML Web

◆ A typical homepage, today:



UNIVERSITY of ABERDEEN

Computing Science

**Alun Preece**  
Senior Lecturer

Email: [apreece@csd.abdn.ac.uk](mailto:apreece@csd.abdn.ac.uk)  
Phone: +44 1224 272291  
Fax: +44 1224 273422

```
<table width="100%" border="0" cellspacing="0"
<tr bgcolor="#CCFFFF">
<td>
<b>Alun Preece</b> <br />
<b>Senior Lecturer</b>
</td>
<div align="right">
Email: <a href="mailto:apreece@csd.abdn.ac.uk">apreece@csd.abdn.ac.uk</a>
Phone: +44 1224 272291 <br />
Fax: +44 1224 273422
</div>
</tr>
</table>
```

A search for "Alun Preece's fax no." gets no help from the HTML markup

UNIVERSITY of ABERDEEN



[Advanced Search](#) [Preferences](#) [Language Tools](#) [Search Tips](#)

Alun Preece fax number

Google Search

Web Images Groups Directory News

Searched the web for **Alun Preece fax number**.

[Ectech mailing list archive: CFP Special Session on Web Mining.](#)

... Next message: **Alun Preece**: "CFP: E ... include: Title of the paper, name, affiliation, postal address, E-mail address, telephone number, and Fax number for each ...  
[www.igec.umbc.edu/ectech/mail/current/0028.html](http://www.igec.umbc.edu/ectech/mail/current/0028.html) - 14k - [Cached](#) - [Similar pages](#)

[Ectech mailing list archive: CFP Special Session on Web Mining.](#)

... Previous message: **Alun Preece**: "CFP: E- Business & the Intelligent Web". ... affiliation, postal address, E-mail address, telephone number, and Fax number for each ...  
[www.igec.umbc.edu/ectech/mail/current/0026.html](http://www.igec.umbc.edu/ectech/mail/current/0026.html) - 14k - [Cached](#) - [Similar pages](#)

[MLnet-list: ECCBR2002 - FINAL deadline is April 5](#)

... scms.rgu.ac.uk Apologies once again for any inconvenience - **Alun Preece** (ECCBR2002 co ... name(s) of the author(s); name, address, phone and fax number, and email ...  
[www.mail-archive.com/community@mlnet.org/msg00367.html](http://www.mail-archive.com/community@mlnet.org/msg00367.html) - 12k - [Cached](#) - [Similar pages](#)

[MLnet-list: ECCBR Deadline Extended by TWO WEEKS](#)

... then please email me at [smc@maths.abdn.ac.uk](mailto:smc@maths.abdn.ac.uk) or **Alun Preece** at [apreece@csd.abdn.ac.uk](mailto:apreece@csd.abdn.ac.uk) ... must include: name(s) of the author(s); name, address, phone and fax number, and email of ...  
[www.mail-archive.com/community@mlnet.org/msg00354.html](http://www.mail-archive.com/community@mlnet.org/msg00354.html) - 12k - [Cached](#) - [Similar pages](#)  
[ More results from [www.mail-archive.com](http://www.mail-archive.com) ]



UNIVERSITY of ABERDEEN

The Semantic Web - Alun Preece

4

## Doesn't XML solve this problem?

- ◆ With XML, custom sets of tags can be defined, which can be:
  - "styled" into conventional HTML
  - processed directly by software - so can be queried...

**Alun Preece**

Email: [apreece@csd.abdn.ac.uk](mailto:apreece@csd.abdn.ac.uk)  
Phone: +44 1224 272291  
Fax: +44 1224 273422

```
<Xfgpds ID="apreece">  
<ghew>Alun Preece</ghew>  
<ngre resource="mailto:apreece@csd.abdn.ac.uk" />  
<hytcx resource="tel:+44-1224-272291" />  
<hytcx resource="fax:+44-1224-273422" />  
</Xfgpds>
```

XML is just syntax - the tags  
don't mean anything!



## The Semantic Web

- ◆ XML alone is not enough - the XML tags need a defined semantics, to make them meaningful
- ◆ Formally, to relate the tag symbols to the things they represent in the real world
- ◆ So the markup becomes a model of the real world



```
<Person ID="apreece">  
<name>Alun Preece</foaf:name>  
<mbox resource="mailto:apreece@c...>  
<phone resource="tel:+44-1224-27...>  
<phone resource="fax:+44-1224-27...>  
</Person>
```

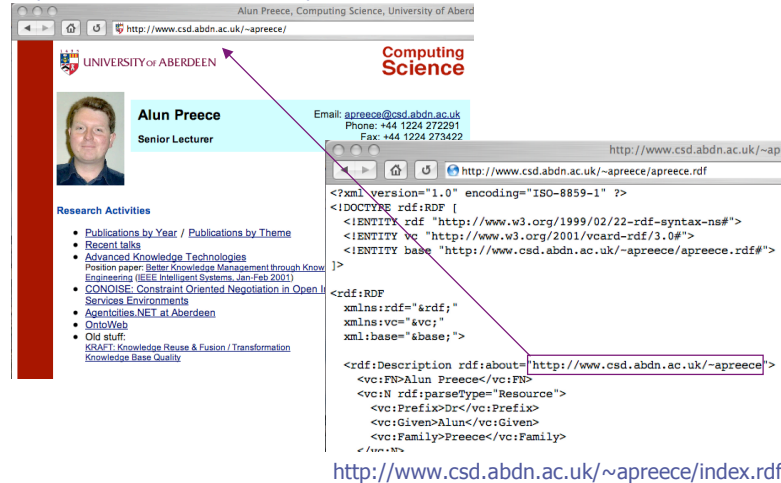


## Two Webs in parallel

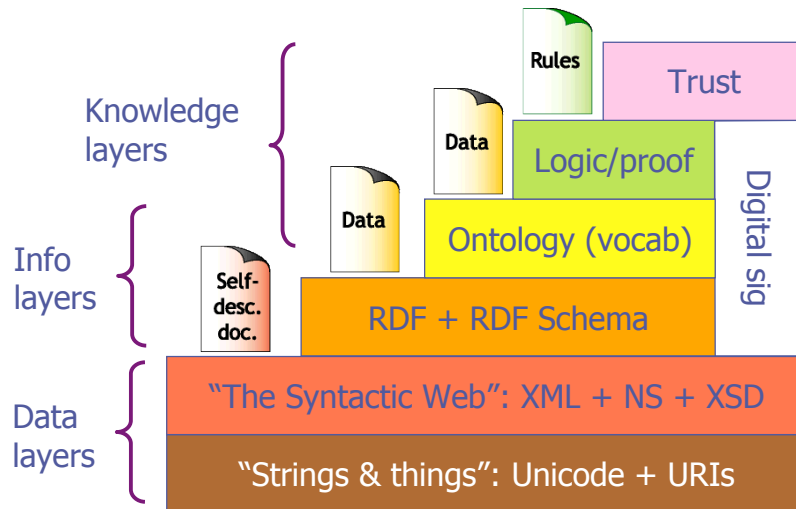
- ◆ The Semantic Web is not a replacement for the current Web
- ◆ Semantic markup is designed to exist alongside HTML markup - often in the form of metadata (data that describes other data)
  - Humans will continue to view the HTML
  - Software can process the Semantic markup
  - Hence the W3C's aim: "better enabling computers and people to work in cooperation"
- ◆ (Because the Semantic markup is in XML, it is possible to generate HTML from it...)

## Web/Semantic Web example

http://www.csd.abdn.ac.uk/~apreece/index.html



```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!DOCTYPE rdf:RDF [
  <ENTITY rdf: "http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <ENTITY vc: "http://www.w3.org/2001/vcard-rdf/3.0#">
  <ENTITY base: "http://www.csd.abdn.ac.uk/~apreece/apreece.rdf#">
]>
<rdf:RDF
  xmlns:rdf="rdf;"
  xmlns:vc="vc;"
  xml:base="base;">
  <rdf:Description rdf:about="http://www.csd.abdn.ac.uk/~apreece">
    <vc:FN>Alun Preece</vc:FN>
    <vc:N rdf:parseType="Resource">
      <vc:Prefix>DK</vc:Prefix>
      <vc:Given>Alun</vc:Given>
      <vc:Family>Preece</vc:Family>
    </vc:N>
  </rdf:Description>
</rdf:RDF>
```



[Adapted from Semantic Web "layer cake" slide due to Tim Berners-Lee]

## Data layers: Unicode, URIs, XML

Like the "traditional" Web, Semantic Web data is based on W3C-recommended standards:

- ◆ Unicode for strings (in all languages)
- ◆ URIs - Uniform Resource Identifiers - to name "things"
- ◆ XML as the standard extensible markup language
- ◆ XML Schema for a variety of primitive datatypes (integer, real number, string, date, URI, ...)
- ◆ XML namespaces to give global scope for tag names

```
<rdf:RDF
  xmlns:vc="http://www.w3.org/2001/vcard-rdf/3.0#"
  <rdf:Description
    rdf:about="http://www.csd.abdn.ac.uk/~apreece">
    <vc:FN>Alun Preece</vc:FN>
    <vc:EMAIL>apreece@csd.abdn.ac.uk</vc:EMAIL> ...
```



- ◆ RDF - **Resource Description Framework** - is the foundation of the Semantic Web standards
- ◆ RDF provides:
  - a simple semantic data model with **classes** (entities) & **properties** (relationships)
  - schema definition constructs (**RDF Schema**) to define simple vocabularies of terms
  - an **XML syntax** for marking-up RDF data
- ◆ RDF is the most well-established and stable aspect of the Semantic Web:
  - many **RDFS vocabularies** are currently available
  - a suite of **software tools** exists to process RDF



- ◆ An RDFS definition for a class called **Person**:

```
<rdf:Class rdf:ID="http://xmlns.com/foaf/0.1/Person" />
```
- ◆ RDFS definitions for three RDF properties - **name**, **mbox** (email address), and **phone**:

```
<rdf:Property rdf:ID="http://xmlns.com/foaf/0.1/name">  
<rdf:range rdf:resource=  
  "http://www.w3.org/2000/01/rdf-schema#Literal" />  
</rdf:Property>  
<rdf:Property rdf:ID="http://xmlns.com/foaf/0.1/mbox" />  
<rdf:Property rdf:ID="http://xmlns.com/foaf/0.1/phone" />
```
- ◆ Note how all these terms are named globally with URIs



## Sample RDF data fragment

```
<rdf:RDF
  xmlns:rdf=
    "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xml:base=
    "http://www.csd.abdn.ac.uk/~apreece/foaf.rdf#">

  <foaf:Person rdf:ID="apreece">
    <foaf:name>Alun Preece</foaf:name>
    <foaf:mbox rdf:resource=
      "mailto:apreece@csd.abdn.ac.uk" />
    <foaf:phone rdf:resource="tel:+44-1224-272291" />
    <foaf:phone rdf:resource="fax:+44-1224-273422" />
    ...
  </foaf:Person>

</rdf:RDF />
```



## Querying RDF

- ◆ RDF statements can be parsed from the XML format into an *RDF model*, for example using Hewlett Packard's *Jena* toolkit
- ◆ Jena RDF models can be queried directly using RDF query languages (initially RDQL; later SPARQL)
- ◆ Example RDQL, "retrieve the phone number(s) of the person whose name is 'Alun Preece'":

```
SELECT ?y
WHERE ( ?x, <foaf:name>, "Alun Preece" )
AND ( ?x, <foaf:phone>, ?y > )
USING foaf FOR <http://xmlns.com/foaf/0.1/>
```



## More than just a smarter Google

- ◆ In addition to offering accurate searching, by querying, Semantic Web data enables many kinds of applications
- ◆ Examples
  - charting communities of friends and colleagues
  - building collaborative community-oriented apps
  - information integration based on standard vocab
  - Web “push”: publish & subscribe
  - automated Web services
- ◆ A lot can often be done with small amounts of semantic markup!



## Six degrees of separation: FOAF

- ◆ The Friend-Of-A-Friend (FOAF) vocabulary covers
  - entities: people, organisations, projects, documents
  - “identifying” details: mbox, homepage, phone, depiction
  - relationships between people: who knows who
- ◆ We’ve already seen FOAF definitions for the Person class, and the name, mbox, & phone properties
- ◆ The FOAF knows property:

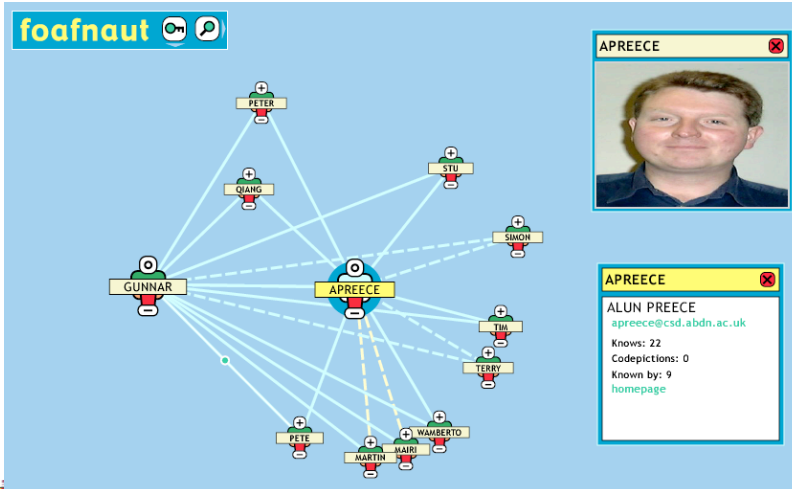
```
<rdf:Property
  rdf:about="http://xmlns.com/foaf/0.1/knows">
  <rdfs:domain
    rdf:resource="http://xmlns.com/foaf/0.1/Person" />
  <rdfs:range
    rdf:resource="http://xmlns.com/foaf/0.1/Person" />
</rdf:Property>
```





## FOAFnaut

- ◆ FOAFnaut is a tool that browses *knows* links:



## The need for ontologies



- ◆ RDF is designed to be simple
- ◆ To define more sophisticated vocabulary, we need to go one layer higher: to the ontology layer
- ◆ The Semantic Web ontology language, OWL, extends RDF with some additional functionality
- ◆ Concrete examples:
  - a **Person** must have at least one **name**
  - a **Person** must have exactly one **age**
  - the class **Person** is the disjoint union of the classes **Man** and **Woman**
  - an email address (**mbox**) belongs to only one **Person**
- ◆ This last example is crucial to FOAF...



- ◆ Core of the World Wide Web Consortium's **Semantic Web** activity
- ◆ In various senses a successor to previous work on "Web-friendly" knowledge modelling languages
  - RDF & RDF Schema
  - DAML-ONT
  - OIL / DAML+OIL
- ◆ W3C's **Web Ontology Working Group** were a "who's who" of the knowledge representation field
- ◆ Released in early 2004 (OWL 1.1 drafted in 2006)



- ◆ **XML**: universal syntax
- ◆ **XML Schema**: defines structure of XML docs
- ◆ **RDF**: datamodel for resource objects
- ◆ **RDF Schema**: basic vocabulary for defining RDF classes & properties, and hierarchies of each
- ◆ **OWL**: extended vocab for defining classes & properties, including
  - **cardinality** (e.g. minCardinality 1)
  - **equality** (e.g. equivalentClass)
  - **relationships between classes** (e.g. disjointWith)
  - **characteristics of properties** (e.g. FunctionalProperty)





- ◆ OWL Lite
  - "RDF-and-a-half"
  - Mainly intended for class hierarchies & simple constraints (cardinality 0 or 1, equality, ...)
- ◆ OWL DL (contains OWL Lite)
  - Description Logic theoretical properties
  - Intended where completeness & decidability are an issue
- ◆ OWL Full (contains OWL DL)
  - Max expressivity; no computational guarantees
  - Supports "Web-scale" & "Web-style" KR&R



## A little ontology goes a long way

- ◆ FOAF uses OWL to define the `mbbox` property:

```
<rdf:Property
  rdf:about="http://xmlns.com/foaf/0.1/mbbox">
  <rdf:type rdf:resource=
    "http://www.w3.org/2002/07/owl#
    InverseFunctionalProperty" />
</rdf:Property>
```
- ◆ This definition means: "mbbox is a personal mailbox, i.e. an Internet mailbox associated with exactly one owner"
- ◆ This means, in database terms, the value of `mbbox` acts as a **primary key** for Persons in the FOAF world - a unique ID



## Ontology mapping

- ◆ OWL can also be used to map one vocabulary to another - hugely important in **information integration** apps
- ◆ Example: the vCard **EMAIL** property is the same as FOAF's **mbox**:

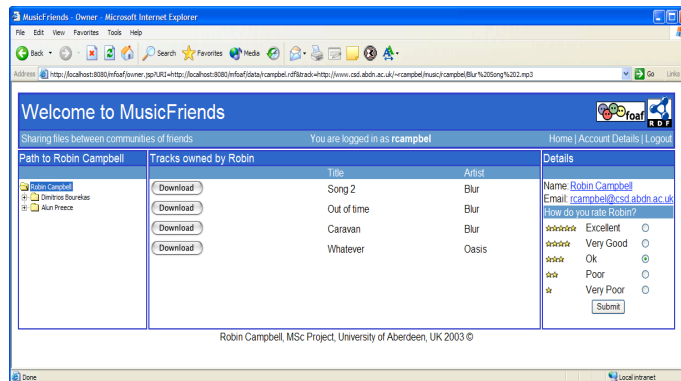
```
<rdf:Property
  rdf:about=
    "http://www.w3.org/2001/vcard-rdf/3.0#EMAIL">
  <owl:equivalentProperty
    rdf:resource="http://xmlns.com/foaf/0.1/mbox" />
</rdf:Property>
```

- ◆ An OWL reasoner could use this equivalence to derive a value for some resource's **vcard:EMAIL** if it can find a value for **foaf:mbox**



## Application: MusicFriends

- ◆ MP3 file sharing among a community of friends
- ◆ Uses FOAF RDF vocab for friend-to-friend links
- ◆ Uses MusicBrainz RDF vocab for MP3 collections



## Application : CS AKTive Space

- ◆ Repository of over 10 million RDF statements
- ◆ Covers entire UK computing science community

## Application: RSS

- ◆ RDF Site Summary (RSS) is an open framework for “publish and subscribe” applications, using RDF
- ◆ Many news sites (and other sites with frequent updates) now provide RSS channels
- ◆ By using a “newsfeed” tool, one can subscribe to RSS channels of one’s choosing
- ◆ When new items are published in RSS/RDF format, subscribers are notified
- ◆ Items can be anything with a URI
  - news stories
  - published documents
  - slides of talks...

## RSS sample

◆ RSS metadata for this tutorial:

```
<rss:item rdf:about="http://www.csd.abdn.ac.uk/~apreece/talks/BCSSGAIBasicsOfAITheSemanticWeb/">
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <rss:title>BCS SGAI Basics of AI: The Semantic Web</rss:title>
  <dc:creator>Alun Preece</dc:creator>
  <rss:description>Slides for the BCS SGAI Basics of AI Tutorial, on the Semantic Web</rss:description>
  <rss:link> http://www.csd.abdn.ac.uk/~apreece/talks/BCSSGAIBasicsOfAITheSemanticWeb/ </rss:link>
</rss:item>
```



## Semantic Web services

- ◆ The key features of the Semantic Web...
    - machine-processable data
    - standard vocabularies
    - compatibility with the “Web family” of standards
  - ◆ ... makes the technology very appealing for automated Web services in all sectors:
    - e-business
    - e-science
    - e-health
    - e-response
    - e-governance
- Web services use Web standards to allow client software to call upon Web servers to carry out tasks - far more than just information retrieval...



- ◆ Semantic Web service: helps a user to schedule a night out in Aberdeen
- ◆ Sources of information, all in RDF:
  - Restaurants (uses standard ontology)
  - Cinema shows (uses standard ontology)
  - Pubs (uses a home-grown ontology)
- ◆ Remembers and recalls user preferences - semantic profiling
- ◆ AI-based scheduler maps RDF data to constraints and produces valid schedules
- ◆ Part of EU-funded [Agentcities.NET](#) project (Worldwide network of intelligent Web services)



- ◆ The Semantic Web is exciting from several perspectives:
  - as a piece of AI / computing science technology
  - as a “new generation” for the Web
  - as a platform for diverse kinds of applications
- ◆ It’s still the Web we know and love:
  - it co-exists with all our messy HTML, etc data
  - it’s a global system: URIs are universal!
  - it’s extremely open
  - it’s not hard to get started...

