The Semantic Web and the PIA project
URL: http://research.nii.ac.jp/~collier/
Nigel Collier, Ai Kawazoe, Tony Mullen, Yoko Mizuta, Tuangthong Wattarujeekrit

World Wide Web today
Computers do not have any ways to process the meaning (semantic contents) of the Web documents. Most of the Web's content is designed for humans to read.

The Semantic Web, the next generation web is being designed so that computer programs can 'understand' the meaning of information in Web resources and carry out sophisticated tasks for users. The 'smart' applications that we hope to see emerge on the Semantic Web include smart browsers, question answering systems, automatic data formatting for different devices in wireless mobile ubiquitous data networks, and support for electronic shopping and appointments through agent software. The Semantic Web is based on Ontology, a knowledge representation defined as "a specification of a conceptualization" (by Gruber 1993).

Question: Who makes the ontology? Who makes the annotation?
Creating the meta-data schemes in the form of ontologies and annotating semantic contents in Web resources are high-cost and time-consuming tasks, and it risks becoming the occupation of a few highly trained professionals unless we can provide support tools, i.e. the "off the shelf software for writing Semantics", and environments that encourage users to participate and share their knowledge.

The Portable Information Access (PIA) project
The PIA project aims to provide:
(1) An environment where domain experts cooperatively construct an ontology and a collection of annotated data based on the ontology
(2) An automatic annotation system based on the machine learning from the annotated data in (1)

We consider that (1) and (2) will reduce the cost and time in constructing metadata and annotated resources. As a component of the system, we have developed Open Ontology Forge, a support tool for ontology creation and document annotation by human experts.
Open Ontology Forge (OOF) and the knowledge model

Open Ontology Forge is a software designed to have integrated functions to support ontology creation, named entity/coreference annotation by human experts. This software provides a convenient environment for human experts to encode:

-- Class-instance relations: the links between deep-level knowledge in the form of ontology and its instances (“the Vertical links” in the figure)
-- Coreference relations (in a broader sense): The links among surface-level instances (various linguistic expressions and images) which refer to the same concept (“the Horizontal links” in the figure)

These relations are important for various purposes, including reasoning with the objects in the document, identifying facts, multimodal access for the information, etc.

Annotation scheme

The central notion with our knowledge annotation scheme is ontology defined in RDF Schema, and annotation is its instance. Annotations of terminologies and coreference relations are saved in a file separated from the original text. Linkage between annotation and texts are inserted as XPointer values.

A key notion with the scheme is the “coreference pool,” which is a set of items which refer to the same concept. This notion plays an important role in encoding coreference relations between multimodal items (ex. texts and images), and ones between items in different documents.
The user interface of OOF

A full Web-browser view of the Web page is maintained in the OOF main window including images.

Ontology Creation

Ontology can be created in a simple manner. Users can define new properties for the ontological classes.

Named Entity annotation

NE annotation can be done in drag-and-drop manner. When a user captures an instance OOF automatically searches for similar instances and offers the user the chance to annotate these too.

Coreference annotation

Drag and drop annotation for coreference: 1) dragging and dropping of an NE to the left pane creates a new coreference pool 2) adding other items into the same pool will encode the coreference relations among them. OOF checks consistency of properties among the members of coreference pool, by unification.
Image Annotation with OOF

Users can select a part of image by clipping it in basic shapes available in Scalable Vector Graphics (SVG).

A drag and drop annotation for images: The consistent procedure for linking images with ontological classes and coreference pools.

Output format

Users can export the result of the annotation to in-text XML format and RDF format.

```
<Bio_515:person rdf:about="bio_515:1000002"
  bio_515:author="aichan"
  bio_515:created="2004-04-13 11:07:01"
  bio_515:expression_type="image"
  bio_515:modified="2004-04-13 11:07:01"
  bio_515:ontology_id="person"
  bio_515:orphan="false"
  bio_515:pool_id="C_000001"
  bio_515:role="true"
  bio_515:svg="<ellipse cx="107" cy="54" ry="64" rx="64" fill="none" stroke-width="2" stroke="#00000000" stroke-opacity="1.0"
   bio_515:term="true"
  bio_515:x:pod="C:\Temp\html\My\news.html#pointer([img(0)])"
  rdf:label="1000002">">
```

Current Status

OOF is now used for constructing annotated corpus from journals in biomedical domain.

Current version of OOF is downloadable from:

http://research.nii.ac.jp/~collier/resources/OOF/index.htm