

The ART Project: An ontology based article preparation tool

Dr Maria Liakata¹ and Dr Larisa N. Soldatova², Department of Computer Science, University of Wales, Aberystwyth
¹mal@aber.ac.uk ²lss@aber.ac.uk

Collaborators: Colin R. Batchelor (RSC), Helen H. Fielding (UCL), Elizabeth Lyon (UKOLN), Stuart Lewis (IS, AU), Claire Q (AU), Simone Teufel (Cambridge), Ross King (AU)

Outcomes of the ART Project

The aim of the ART project was to create a tool to allow the annotation of **core scientific concepts** in papers. To this effect we developed a tool (**SAPIENT**) to enable users (authors, editors, reviewers, experts) to annotate papers with concepts such as 'Goal', 'Motivation', 'Object', 'Hypothesis', 'Background', 'Model', 'Experiment', 'Method', 'Observation', 'Result', 'Conclusion'.

These concepts constitute the **CISP meta-data** and were verified through an on-line survey addressed to researchers. The CISP meta-data were accompanied by a set of guidelines for their implementation as an annotation scheme.

We worked with 16 experts, who used the guidelines and SAPIENT to create a corpus of 225 papers manually annotated with CISP concepts (**ART corpus** > 1 million words). These papers cover topics in Physical Chemistry and Biochemistry and were provided by the Royal Society of Chemistry (RSC). However, CISP and SAPIENT are domain independent and our approach can be applied to a wide range of scientific disciplines. Our project outcomes will facilitate information extraction from papers and constitute a milestone towards the automatic refereeing of papers.

The CISP meta-data:

Background Conclusion Experiment Goal Hypothesis Method Model Motivation Object Observation Results

SAPIENT:

[Index](#) | [Refresh](#) | [Auto Annotate](#) | [Clear Auto Annotations](#) | [Clear Own Annotations](#) | [Save](#) | [Help](#)

Background Conclusion Experiment Goal Hypothesis Method Model Motivation Object Observation Results

1 A new approach for **DNA** detection by **SERRS**

Object New Obj1

2 A new approach for the detection of **DNA** using surface enhance resonance Raman **scattering** (SERRS) is reported.

Object New Obj1

3 The majority of existing techniques use **fluorescence spectroscopy** with advanced probe design to provide information on the identity of specific **DNA** sequences down to single **base** resolution.

Method Old Met1

4 A new approach to the labelling of **DNA** is discussed which uses **Michael addition** to couple **thiolated DNA dye** labels specifically designed to attach to **silver** surfaces.

Method New Met2

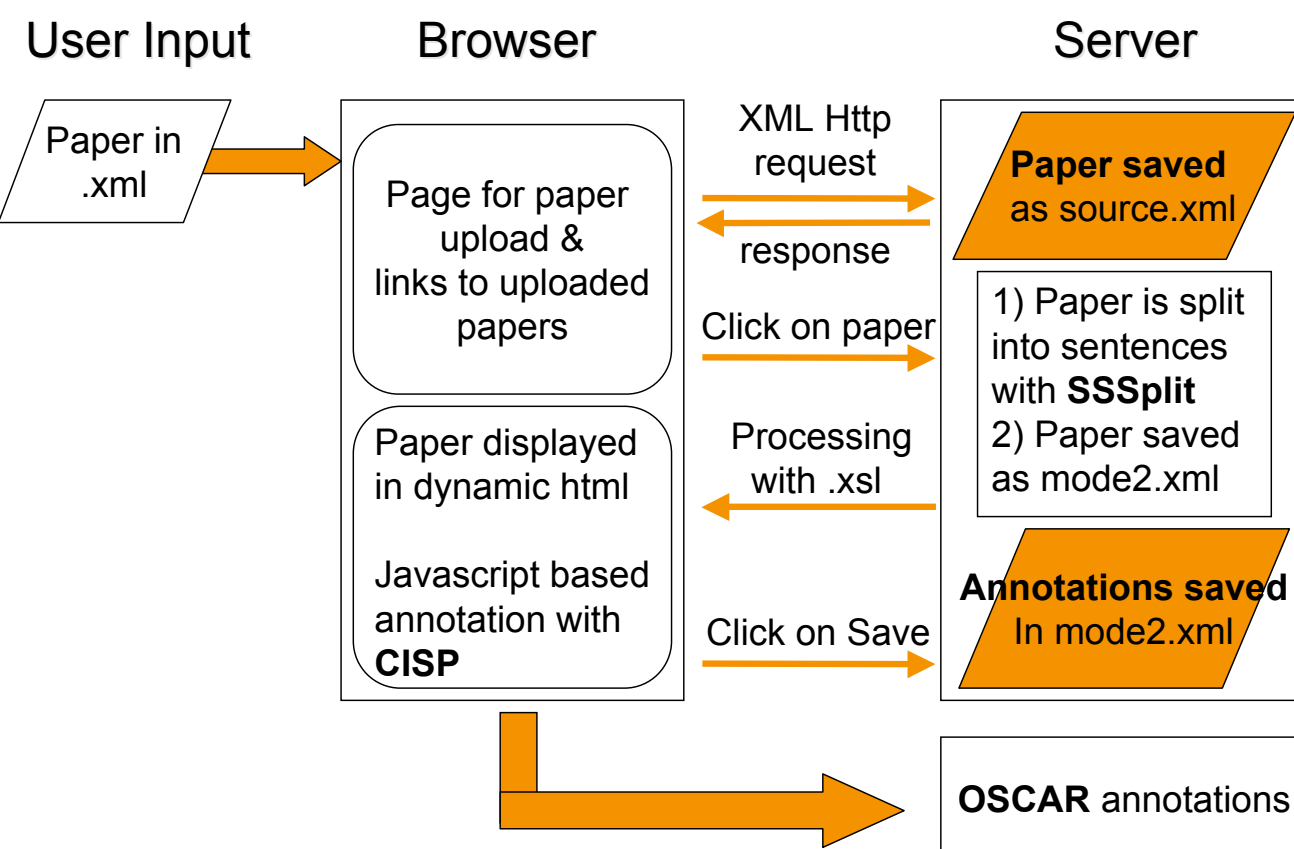
5 When combined with existing **fluorescence** detection of **DNA** using commercially available labels new class of biomolecular probe known as a SERRS Beacon was produced.

Object New Obj1

6 The detection techniques of **fluorescence** and surface enhanced resonance Raman **scattering** (SERRS) are combined to give a sensitive and selective system for use in the development and creation of novel assays for specifically defined targets.

Conclusion None Con1

SAPIENT Architecture



Download from: <http://www.aber.ac.uk/compsci/Research/bio/art/sapient/>

The ART corpus:

Specific technique allowing sharp fingerprints spectra to be obtained, thus allowing discrimination between components in a sample without the need for separation...
 In addition, it is an extremely sensitive technique with single molecule detection reported-REF TYPE="P" text="6,7" ID="cit6 cit7">6,7</REF>...
 The work reported here was carried out using citrate reduced silver nanoparticles since, by careful ci

The **ART corpus** was developed primarily to add value to scientific papers, through semantic markup that would make it easier for natural language processing and semantic web applications to automatically extract information pertaining to core scientific concepts.

The ART corpus can also be used as a training set for machine learning algorithms, in order to automate the annotation of papers with CISP meta-data. The sustainability of and the benefits obtained from annotating papers with CISP meta-data will be investigated by the JISC funded SAPIENT Automation (SAPI-ENTA) project.

Download from: <http://cadair.aber.ac.uk/dspace/handle/2160/1979>