Bio-ontologies SIG/ISMB'07

July 20, 2007

ART: An ontology based tool for the translation of papers into Semantic Web format

Larisa N. Soldatova[1], Colin R. Batchelor[2], Maria Liakata[1], Helen H. Fielding[3], Stuart Lewis[1] and Ross D. King.[1]

1) University of Wales, Aberystwyth, UK;
2) Royal Society of Chemistry Publishing, UK;
3) University College London, UK .

Plan of the presentation

- Introduction to the ART;
- Related projects:
  - The Prospect project;
  - An ontology of experiments EXPO;
  - The ROAD project;
  - eBank.
- The ART project:
  - The ART system;
  - Examples.
- Discussion: An ontology of science.

The ART project (an ontology based ARticle preparation Tool)

- Translating scientific papers into a format with an explicit semantics.
- Explicit linking of repository papers to data and metadata.
- Creation of an example ‘intelligent’ digital repository.

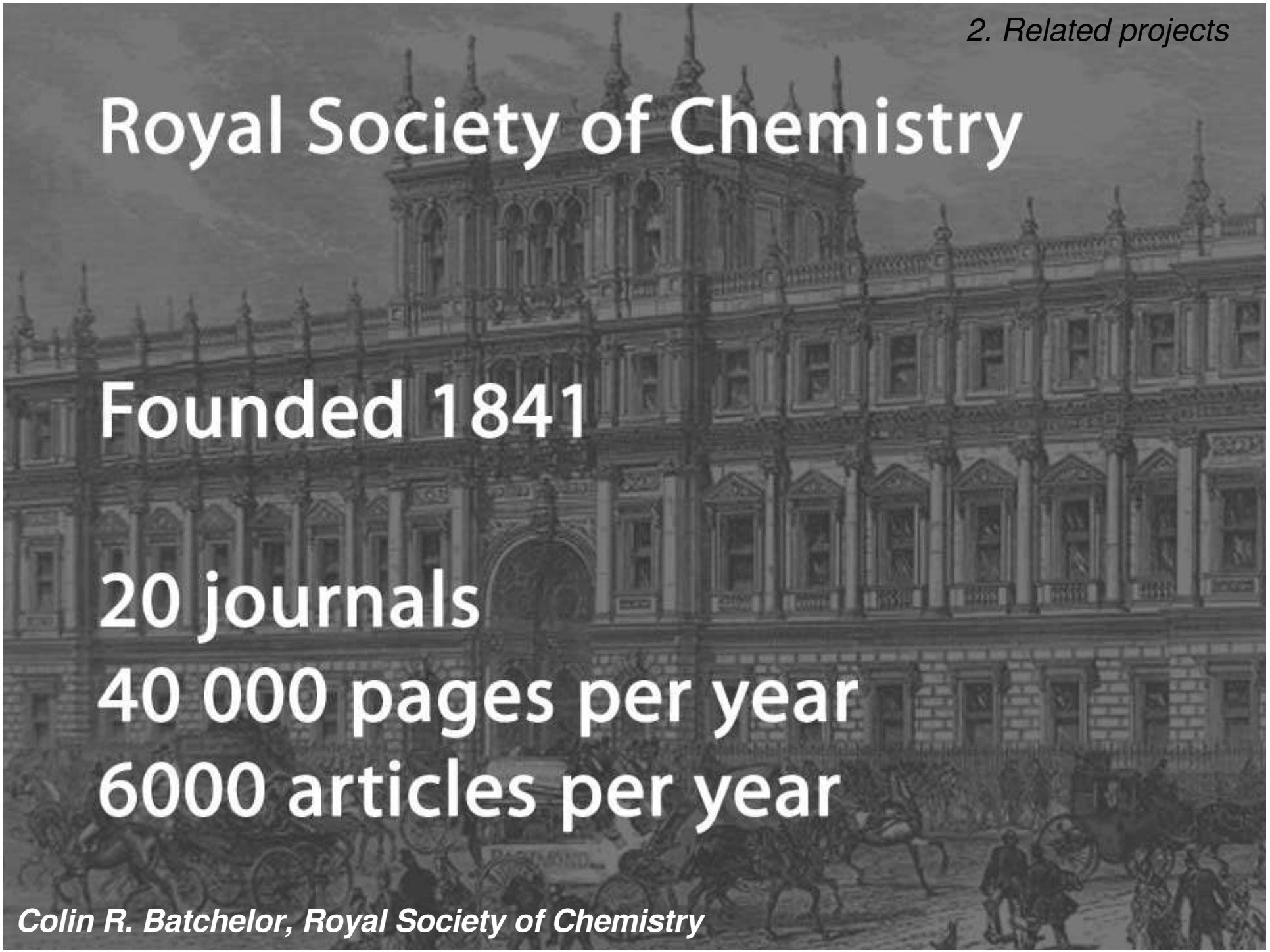
http://www.aber.ac.uk/compsci/Research/bio/art/

Motivation:

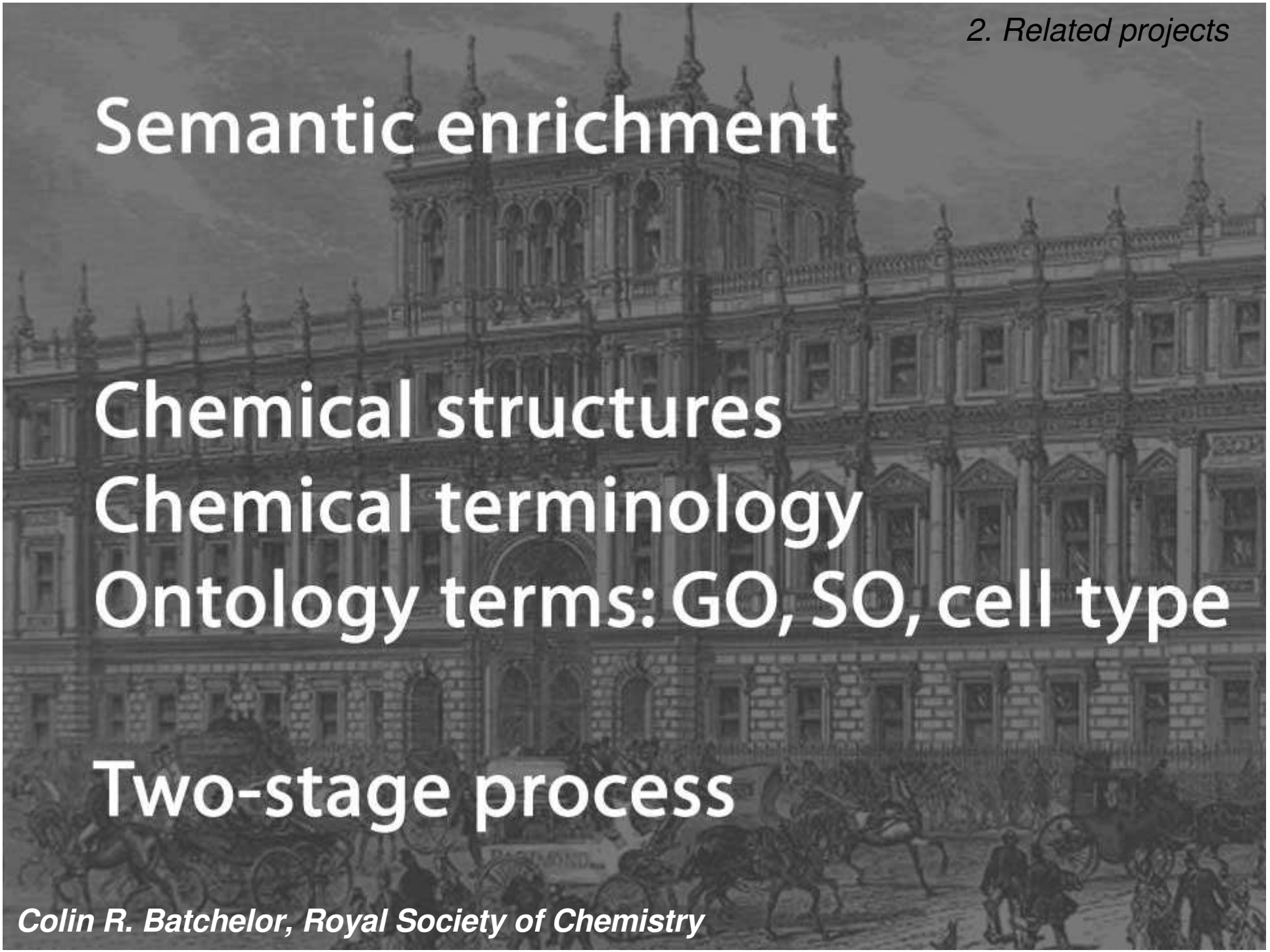
- to improve information access;
- to provide semantic clarity and explicitness of represented information and knowledge;
- to promote the sharing of research results;
- to facilitate text mining and knowledge discovery applications.

www.ProjectProspect.org

Colin R. Batchelor, Royal Society of Chemistry



Colin R. Batchelor, Royal Society of Chemistry



Colin R. Batchelor, Royal Society of Chemistry

Industry Track: 04
Project Prospect: Life Science Informatics at Royal Society of Chemistry Publishing

Sunday, July 22
4:50p.m. - 5:15 p.m.
Room: Hall M

Presented by: Colin R. Batchelor, Royal Society of Chemistry

HTML

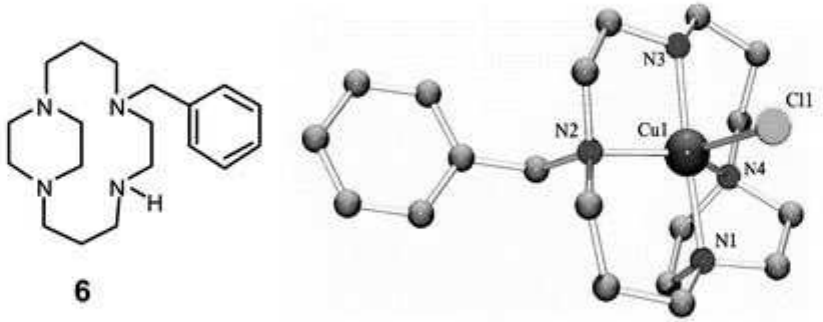


Fig. 1 Ball and stick representation of the X-ray crystal structure of the copper(II) complex formed with the benzyl derivative of the piperazine tetra-aza macrocycle. H-atoms and the $[CuCl_2]^-$ counter anion have been omitted for clarity.

Paramagnetic metal ions affect the fluorescence properties of their ligands[23] and fluorescence quenching has been used to [illegible] magnitude of the effect is expected to be reduced by an increased distance between the fluorophore and the chelated metal i[illegible]he fluorescence (approx. 75%) was observed on formation of $Cu\mathbf{5}Cl_2$; however, the fluorescence emission was still sufficient to [illegible]ogical systems.† Complex stability is also important. The copper(II) complex of a related piperazino cyclam macrocycle has been s[illegible]vorkers and shown to have a demetallation half life of >20 days in 1 M $HClO_4$.[25]

Cellular binding assays were carried out to determine whether the mono-macrocyclic fluorescent compounds interact wit[illegible]eptor. A T-lymphocyte cell line (Jurkat) that expresses high levels of the CXCR4 receptor was used. Compounds **5** and $Cu\mathbf{5}Cl_2$ were analysed for binding to the Jurkat cells by flow cytometry in competition with specific anti-CXCR4 monoclonal antibodies.

Two CXCR4-specific, monoclonal antibodies were used (mAbs 44716.111 and 44717.111), both of which bind to similar regions of the extracellular loop.[18] The binding assay was carried out by saturation of the receptors with a high concentration of compound (*ca.* 25 μM) followed by washing and introduction of the specific antibodies. The amount of antibody present on the cell surface was then quantified by addition of a secondary anti-mouse IgG fluoresceinated reagent by flow cytometry. The free macrocycle conjugate showed no shift from the positive control, suggesting that either no binding occurs or **5** is completely displaced by the antibodies [Fig. 2(A)]. The metal complex $Cu\mathbf{5}Cl_2$ binds and competes successfully with both antibodies showing a significant shift from the positive control [Fig. 2(B)].

Colin R. Batchelor, Royal Society of Chemistry

Ontology of scientific experiments EXPO

EXPO* v.1

http://sourceforge.net/projects/expo

2006 nomination for World Technology Award (software).

Soldatova, LN & King, RD (2006) An Ontology of Scientific Experiments. Journal of the Royal Society Interface **3/11: 795-803.*

eBank

Aims to provide a technological solution to the access and curation of digital resources:

http://www.ukoln.ac.uk/projects/ebank-uk/

The project is being led by UKOLN in partnership with the Intelligence, Agents & Multimedia Group, Department of Electronics & Computer Science, and the Department of Chemistry, University of Southampton and the Digital Curation Centre:

http://www.dcc.ac.uk/

ROAD (Robot-generated Open Access Data) project

http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/road.aspx

ART as an authoring tool

- ***Assist:*** help the user to identify the high level concepts about an investigation in the text. (i.e. hypothesis, object of investigation, conclusion) and domain ontology entities.
- ***Explain:*** provide examples and explanations if necessary.
- ***Represent:*** help the author to represent his/her research results directly in OWL format. The system will ask to input required metadata about the research.

Support for ontologies

- **OBI** (Ontology for Biomedical Investigations)
  i.e. <investigation>, <objective>, <conclusion>.
- **EXPO** (Ontology of scientific experiments)
  i.e. <hypothesis>, <method>, <result>.
- An **ART**-extension
  i.e. <analytical method>, <model>, <equation>.
- Domain ontologies
  - **ChEBI** (Chemical Entities of Biological Interest);
  - **REX** (ontology of physico-chemical methods and properties);
  - **FIX** (ontology of physico-chemical processes);
  - An **ART-d**omain extension.

an ontology for the ART system (a fragment):

- Investigation component
  - Hypothesis of investigation
  - Experiment
  - Goal of investigation
  - Object of investigation
  - Result of investigation
  - Model
  - Technology
  - Motivation
  - Experiment component
  - Problem analysis
  - Example
  - Method of investigation
    - Analytical method
    - Observational method
    - Experimental method
- Linguistic component
  - Paper section
    - Title
    - Abstract
    - Introduction
    - Experimental
    - Results and discussion
    - Conclusion
  - Paper
  - Representation

ART as a translator

- ***Annotate:*** ontology-based markup of the input text;
- ***Recognize:*** automatically find a concept/ synonym in the text;
- ***Get feedback:*** from a user;
- ***Link:*** concept/ synonym in the text to the external source;
- ***Generate***: summary; RSS feed.

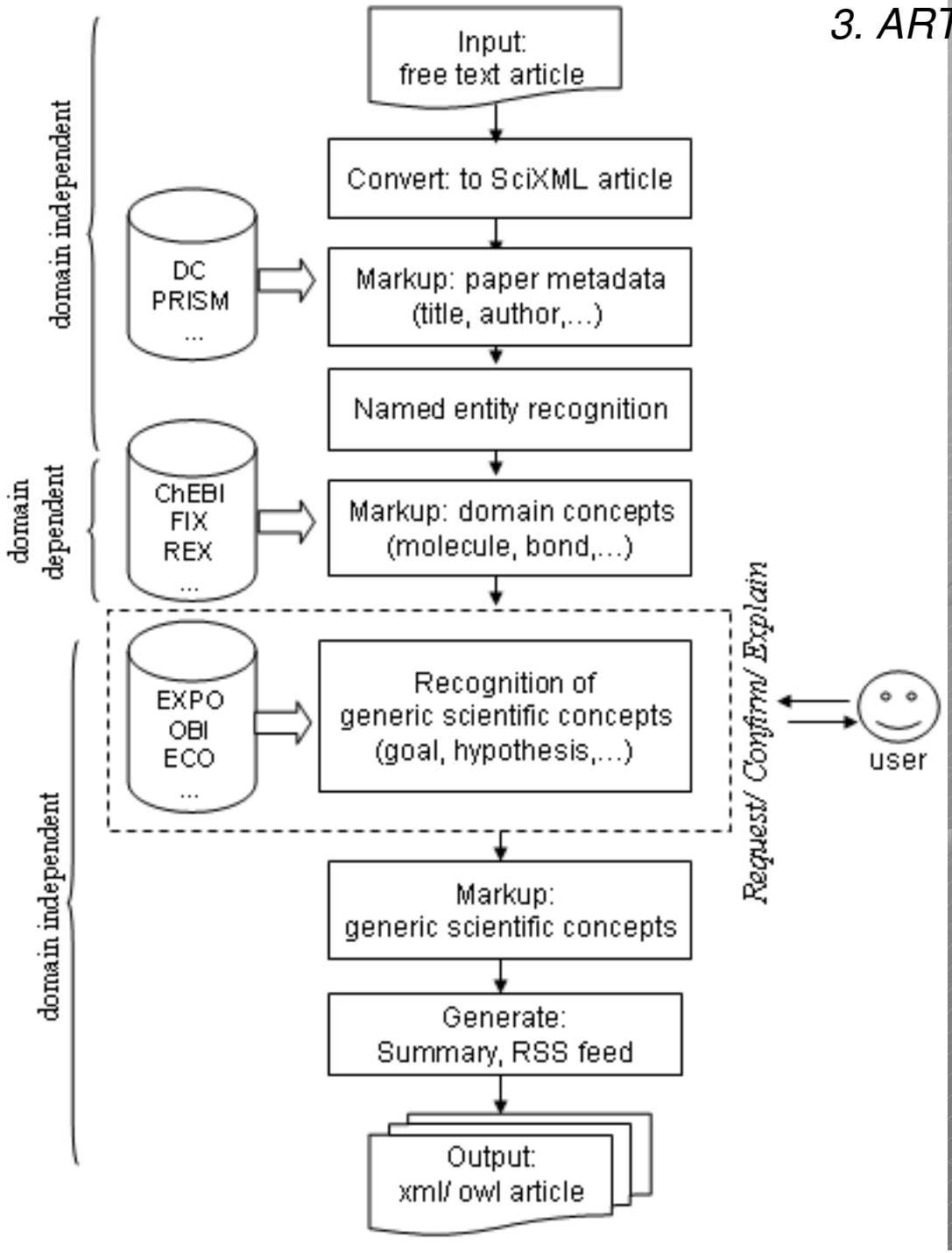
Methodology

- ‘supervised search’;
- machine learning:
  - concepts recognition;
  - features: ‘zoning’ (i.e. “aim”, “own”); meta-discourse markers (i.e. “this paper is aimed at…”);
- interactive learning;
- evaluation of the system: survey of generated summaries; text mining application.

Concepts for automatic recognition:

<motivation for investigation>
<goal of investigation>
<object of investigation>
<hypothesis>
<method>
<model>
<experiment>
<observation>
<result>
<conclusion>

A flow diagram for the ART system



A fragment of an annotated article

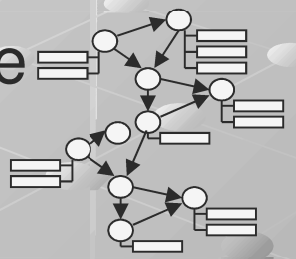
<EXPO: investigation> *map to*
<DC: title> *map to*
<OBI: investigation>
The physical origin of large covalent-ionic resonance energies in some two-electron bonds.
<EXPO: goal>
"studying in detail all the aspects of bond formation in a series of molecules that each display a range of bonding features: H2 and C2H6 as members of the classical family of covalent bonds, Cl2 as a bond exhibiting significant CS character, and the series N2H4, H2O2, and F2 as molecules exhibiting increasing CS character from left to right of the periodic table."
<EXPO: object of investigation> *map to*
<OBI: investigation object role>
characteristics of CS bond in H2, C2H6, Cl2, N2H4, H2O2, F2 molecules
<EXPO: method>
valence bond calculation on two levels:
valence-bond-self consistent field (VBSCF)
breathing-orbital valence bond (BOVB)

Ontology of science

e-Science

Ontology of science

(formalization of scientific methods, technologies, infrastructure of science)



- Controlled vocabulary for science;
- Formalized computational representation of scientific investigations;
- Unified standard for representation, annotation, storage, and access to experimental results;
- Automated reasoning over experimental data and conclusions.

Summary

- The ART tool;
- An ontology support;
- An example repository of papers in OWL.

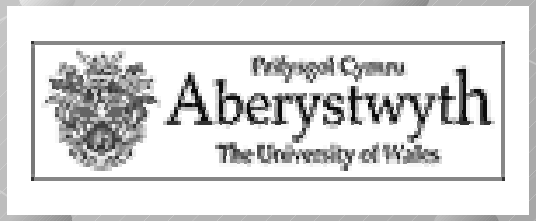
Acknowledgements

Joint Information Systems Committee, UK

Research councils, UK

Biotechnology and Biological Sciences Research Council, UK

University of Wales, Aberystwyth, UK



Osaka University, Japan

Thank you