

Semantic Enrichment of the Scientific Literature, 30th – 31st March 2009

We would like to invite you to the EBI workshop on Semantic Enrichment of the Scientific Literature which is co-organised and sponsored by the BootStrep project (contact Dietrich Rebholz-Schuhmann, rebholz@ebi.ac.uk) and the EBI Industry Programme (contact Dominic Clark, clark@ebi.ac.uk).

The workshop venue is the Wellcome Trust Conference Centre, Hinxton, Cambridge, UK.

Motivation

Over the last 10 years, innovation has changed the ways in which scientific publications are gathered and delivered to the public. Since the start of the electronic era:

- publishers have moved from paper presentation of their content to electronic delivery;
- the US National Library of Medicine (NLM) has opened up its archives and delivered Medline abstracts to the public in electronic form;
- open access publishers have been making their content freely available;
- curation teams are increasingly working with the publishers to gather ever more data and provide it to the public and
- proposals have been made that authors should contribute more details to their manuscript (FEBS letter experiment).

These changes require novel ways to capture and deliver the content to the public and to exchange the content and the annotations between different sites. This has led to increased activities to capture more information from the authors, to align it with the bioinformatics data resources, to deliver the content as part of the scientific literature and to improve the interoperability between existing automatic systems for text processing and exploitation.

This workshop will focus on semantic enrichment of the scientific literature. To this end, workshop participants will have the opportunity to hear about and discuss solutions that capture information from the authors directly and that deliver documents with their annotations. Furthermore, we will discuss the needs of different user groups for the benefits from the scientific literature of the future, e.g. information scientists, researchers, automatic text processing and data mining research community, ontologists, others.

Participants

The participants in this workshop are all users who profit from better information retrieval (e.g., librarians and information scientists supporting industrial researchers) and information provision (e.g., bioinformatics research community). In addition, members of the text mining research community, members of publishing companies and industrial users of scientific information, curation teams and teams working on ontological or terminological resources.

Intended Outcomes

The intended outcomes of the workshop are:

- to exchange views on the reuse of literature in all possible ways and to the benefit of all involved parties;
- to have a better understanding for the infrastructure requirements coming from the automatic gathering and distribution of semantically enriched text (open standards and connectivity);
- to exchange views on what contributions could come from the publishers with regards to better exploitation of the publicly available literature and
- to assess current solutions for the gathering of semantic details from the authors while writing their scientific manuscripts.

Programme Overview

The workshop programme has been divided into three separate days – each with its own focus. The foci are as follows:

- March 30th: **Reliable factual data from the literature based on ontological resources** (open meeting, Francis Crick Auditorium). Presentations and discussions on advanced solutions to model ontological resources for gathering facts from the literature and for integration into a fact database: gene regulatory events as a working example.
- March 31st: **Semantic Enrichment of the literature for the benefit of all users** (open meeting, Francis Crick Auditorium). Existing solutions for the processing and standardization of annotations in the scientific literature: authoring solutions, access to data through publishers, requirements from curators and “information scientists” in pharmaceutical and other industrial companies.

Acknowledgement

We are grateful to Ian Dix (AstraZeneca), Ian Harrow (Pfizer) and Wendy Filsell (Unilever) for their advice in the construction of the workshop programme. This workshop is sponsored by the EC STREP project ‘BootStrep’ (FP6-028099, www.bootstrep.org) and by the industry program at the EBI (<http://www.ebi.ac.uk/industry/ind-prog-index.html>).

Detailed Programme

March 30th: Reliable factual data from the literature based on ontological resources (open meeting, Francis Crick Auditorium).

Presentations and discussions on advanced solutions to model ontological resources for gathering facts from the literature and for integration into a fact database: gene regulatory events as a working example.

09.00	Registration Open (tea/coffee)
10.30	Welcome and Introductions (Dietrich Rebholz-Schuhmann)
	Session 1: Semantic representation of Gene Regulatory Events (GREs) Confirmed Speakers (exact times and titles to be agreed). <ul style="list-style-type: none"> • “Ontologies and text mining” (Junichi Tsujii, University of Tokyo) • “Gene regulation ontology: design and exploitation for information extraction” (Jung-Jae Kim, EBI, Hinxton, Cambridge, U.K.) • “OregAnnO: curated gene regulatory events” (Stephen Montgomery, Sanger, Wellcome Trust Genome Campus, U.K.)
12.30	Lunch
13.45	Session 2: Identification of Gene Regulation Events in the scientific literature Confirmed Speakers (exact times and titles to be agreed) <ul style="list-style-type: none"> • “BioLexicon” (Simonetta Montemagni, CNR, Pisa) • “Identification of gene regulatory events from the literature” (NN, Friedrich-Schiller University, Jena, Germany) • “Language Resource Assessment for Information Access”, Su Jian, InfoComm Research, Singapore
15.30	Tea/Coffee
16.00	Session 3: Clinical data and Novel publishing solutions Confirmed Speakers (exact times and titles to be agreed) <ul style="list-style-type: none"> • “Ontology development for information extraction (ODIE) from clinical text” (Wendy Chapman, University of Colorado) • “Embedding semantic data during manuscript authoring” (Lynn Flinn, University of California San Diego) • “PaperMaker: consistency analysis of published manuscripts”, Piotr Pezik, EBI, Hinxton, Cambridge, U.K.
17.30	Closing remark and Discussion (Dietrich Rebholz-Schuhmann)
18.00	Pre-dinner drinks and networking
19.30	Workshop Dinner (Hinxton Hall restaurant)

March 31st: Semantic Enrichment of the literature for the benefit of all users

(open meeting, Francis Crick Auditorium).

Existing solutions for the processing and standardization of annotations in the scientific literature: authoring solutions, access to data through publishers, requirements from curators and “information scientists” in pharmaceutical and other industrial companies.

08.30	Registration Open (tea/coffee)
09.30	Welcome and Introductions (Dietrich Rebholz-Schuhmann & Dominic Clark)
10.00	Session 1: User’s and publishers perspective of future publishing Confirmed Speakers (exact times and titles to be agreed). <ul style="list-style-type: none"> • Keynote: “Semantic Enrichment of Elsevier’s literature”, Anita de Waard, Elsevier, Amsterdam, NI • “The needs of information scientists in Pharma companies” (Jasen Chooramun, AstraZeneca)
11.00	Tea/Coffee
11.30	Session 1 cont: Confirmed Speakers (exact times and titles to be agreed) <ul style="list-style-type: none"> • The needs of bioinformaticians in Pharma companies (Ian Harrow, Phoebe Roberts, Pfizer) • “Elixir: European Infrastructure to support Interoperability between Text Repositories and Biological Databases” (Alfonso Valencia, CNIO, Madrid, Spain)
12.30	Lunch
13.45	Session 2: Semantic enrichment Confirmed Speakers (exact times and titles to be agreed) <ul style="list-style-type: none"> • Keynote: “Completing the Circuit: Publishing with Linked Data” (Eric Neumann, Clinical Semantics Group) • “WikiGene: Novel ways to publish information on the Web” (Robert Hoffmann, MIT) • “CALBC + UKPMC: Producing a large scale annotated corpus for standardized literature” (Dietrich Rebholz-Schuhmann, EBI)
15.30	Tea/Coffee
16.00	Session 3: new solutions for publishing Confirmed Speakers (exact times and titles to be agreed) <ul style="list-style-type: none"> • Keynote: “Reconciling annotations made to different terminologies: The role of terminology integration services” (Olivier Bodenreider, NLM) • “Solutions for novel text mining” (Sophia Ananiadou, NaCTeM, Manchester) • “Integration of ontological resource into novel information extraction solutions” (Udo Hahn, FSU)
17.30	Closing remark and Discussion (Dietrich Rebholz-Schuhmann)