

Community Pathway Curation: Tools and Technologies

Name of the speakers:

Yukiko Matsuoka^{1,2}, Hiroaki Kitano^{1,3} Brian Kemper⁴, Sophia Ananiadou^{5,6}

1 The Systems Biology Institute, 2 JST ERATO Kawaoka Infection-induced host response project. 3 Okinawa Institute of Science and Technology, 4 University of Tokyo, 5 University of Manchester, 6 National Centre for Text Mining

Purpose and background of the subject area:

Curation is essential to creating an accurate pathway model. Modeling an extensive gene-regulatory and biochemical pathway requires extensive data and literature mining. Yet once curated models need constant revision to be up-to-date.. What is needed is a framework to facilitate tracking and an update mechanism for modelers and researchers in the community to collect the latest information, to find the evidence from literature and to contribute to the collaborated model building, sharing, comparing and curation.

To provide such a curation framework, we developed the system “Payao” (available <http://www.payaologue.org>), a web-based, collaborative, biological knowledge aggregating system. It enables the community to work on the same models simultaneously, insert annotation tags, exchange comments, record the discussions and eventually update the models. Payao adopts the community standards, accepts Systems Biology Markup Language (SBML, <http://sbml.org>) format, displayed in Systems Biology Graphical Notation (SBGN, <http://sbgn.org>) compliant CellDesigner graphical notation, and annotations conforming to Minimal Information Required In the Annotation of Models (MIRIAM <http://www.ebi.ac.uk/miriam/>).

PathText (<http://www.pathtext.org>) is an integrated environment for combining standards compliant with biological pathway models and original papers relevant to selected parts of the pathway, through the use of text mining technology and tools to facilitate the creation of manual annotations. PathText distinguishes itself from other pathway editing tools by providing a seamless combination of advanced text mining technologies. These include (i) MEDIE (<http://www-tsujii.is.s.u-tokyo.ac.jp/medie/search.cgi>) an intelligent search engine designed to retrieve biomedical facts and events from the entire MEDLINE, based on indexing by natural language processing and text mining techniques including deep syntactic analysis of individual sentences; (ii) KLEIO (<http://www.nactem.ac.uk/software/kleio/>), a semantic information retrieval system based on named entities (e.g. proteins, genes, metabolites, etc) and (iii) FACTA+ (<http://www.nactem.ac.uk./software/facta>) an interactive search system that finds direct and indirect associations between entities.

PathText provides a user friendly interface by which one can freely move from pathways to relevant articles, or to parts of full papers annotated by biologists during the model construction phase. The central component of PathText is the Integrator, which coordinates interaction between (1) Payao Pathway Visualizations, (2) Text Mining Systems, (3) Manual Annotations and (4) the PathText Results GUI.

In this tutorial, we will explain general curation workflow and tools and technologies available today. We will also have hands-on to go through the actual curation process. The participants will get a hands-on experience of using text mining services.

Requirements: The participants of the tutorial are invited to bring their own laptops to get hands-on experience for “mini-mapathon”.

Tutorial Outline (Approximate Length: 4hours)

1. Community Pathway Curation Overview (2h.)

- Curation tools and technologies (30 min.)
- Text Mining (=TM) technology for Curation (PathText and TM services: 45 min.)
- Community Curation Platform: Payao (45 min.)

2. Mini Mapathon: (1.5h)

Participants would curate a sample signaling pathway model using the TM and curation platform PAYAO.

3. Discussion & Q &A (0.5h)

Required facilities: Standard AV system, internet connection.

The participants **need to bring their own laptop.**

Internet access is required for speakers and/or participants.