TEXT AND KNOWLEDGE MINING FOR PHARMACOGENOMICS: GENOTYPE-PHENOTYPE-DRUG RELATIONSHIPS

K. BRETONNEL COHEN
Department of Pharmacology, University of Colorado School of Medicine
Aurora, Colorado, USA
Email: kevin.cohen@gmail.com

YAEL GARTEN
Biomedical Informatics Program and Department of Genetics, Stanford University
Stanford, California, USA
Email: ygarten@stanford.edu

NIGAM SHAH
Stanford Center for Biomedical Informatics Research, Stanford University
Stanford, California, USA
Email: nigam@stanford.edu

UDO HAHN
Jena University Language and Engineering Laboratory, Jena University
Jena, Germany
Email: Udo.Hahn@uni-jena.de

Text mining for pharmacogenomics is poised to become a major research topic. This session at the Pacific Symposium for Biocomputing is meant to be a catalyst to push it from potential to realization.

Many research areas are relevant to the general topic of text mining for pharmacogenomics. Minimally, the field calls for work on the following:

- Named entities relevant for the pharmacogenomics literature. This includes drugs, chemicals, diseases, genes, and genomic variations.
- Relations relevant for the pharmacogenomics literature. This includes a wide range of relation types, ranging from simple co-occurrence to is-effective-for, treatment-of, is-associated-with, etc. It is not clear what the full set of relations is, and this presents a challenge for knowledge engineering.
- Infrastructure relevant for mining the pharmacogenomics literature. This includes corpora and gold standards, such as PharmGKB, i2b2 data, the Comparative Toxicogenomics Database, and the Arizona Disease Corpus.
- Evaluation of the state of the art, including both novel evaluations and novel approaches to evaluation.
- Reasoning systems applied over resources like the PharmGKB knowledge base.

The session received a number of relevant submissions, on topics ranging from ontologies; phenotype studies and animal models; drug repurposing; mutation combinations and drug resistance; terminologies; drug-drug interactions; reasoning over syntax and semantics; ranking gene-drug interactions; and genetic variants that affect drug response. A rigorous review process resulted in the selection of a set of papers that arguably represents the best of the state of the art in text and knowledge mining for pharmacogenomics, focusing on genotype-phenotype-drug relationships. Much work remains to be done in the area; we hope that this represents just the beginning of a new flowering in studies of text mining for pharmacogenomics.