BMC Bioinformatics

ORAL PRESENTATION

Open Access

Integrating text mining into high-throughput assay analysis

K Bretonnel Cohen

From Workshop on Advances in Bio Text Mining
Ghent, Belgium. 10-11 May 2010

There are two basic paradigms of use cases for the output of text mining tools in the genomics subfield of biomedical natural language processing (BioNLP). In the most common one, the tool is designed to produce output that will be viewed by an individual researcher, most commonly a database curator or a bench scientist. This use case has been heavily studied, thanks in part to shared tasks like BioCreative and TREC Genomics. The other use case is the integration of text mining into high-throughput assay analysis. This includes tasks ranging from sequence alignment to the evaluation of gene expression array data. In the past, text mining has been applied to these areas either as a post-processing step, or as an integrated part of the analysis algorithm. More recently, our lab has developed Hanalyzer, a 3R tool—a tool for the knowledge-based analysis of high-throughput assays based on Reading, Reasoning, and Reporting about experimental results in the context of pre-existing knowledge, including knowledge from text mining. The tool is designed to facilitate exploration of experimental data, explanation of observed patterns in the light of what is already known about the entities involved, and the generation of novel hypotheses. Results from a study of craniofacial development demonstrate that the system can be used to explain patterns in gene expression and to generate a set of hypotheses about the roles of four genes previously not known to be involved in tongue development. These hypotheses were experimentally validated by in situ hybridization and may have clinical consequences related to cleft lip and palate.

Published: 6 October 2010

Correspondence: kevin.cohen@gmail.com
Center for Computational Pharmacology and The MITRE Corporation, USA

Submit your next manuscript to BioMed Central and take full advantage of:

• **Convenient online submission**
• **Thorough peer review**
• **No space constraints or color figure charges**
• **Immediate publication on acceptance**
• **Inclusion in PubMed, CAS, Scopus and Google Scholar**
• **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central