

**Workshop on Linking Natural  
Language Processing and Biology:  
Towards Deeper Biological Literature  
Analysis 2006**

**(BioNLP 06)**

**Held at HLT/NAACL 2006**

**New York, New York, USA  
8 June 2006**

**ISBN: 978-1-61738-492-9**

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571  
[www.proceedings.com](http://www.proceedings.com)

**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2006) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2010)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

## Table of Contents

<i>The Semantics of a Definiendum Constrains both the Lexical Semantics and the Lexicosyntactic Patterns in the Definiens</i> Hong Yu and Ying Wei .....	1
<i>Ontology-Based Natural Language Query Processing for the Biological Domain</i> Jisheng Liang, Thien Nguyen, Krzysztof Koperski and Giovanni Marchisio .....	9
<i>Term Generalization and Synonym Resolution for Biological Abstracts: Using the Gene Ontology for Sub-cellular Localization Prediction</i> Alona Fyshe and Duane Szafron .....	17
<i>Integrating Ontological Knowledge and Textual Evidence in Estimating Gene and Gene Product Similarity</i> Antonio Sanfilippo Christian Posse, Banu Gopalan, Stephen Tratz and Michelle Gregory .....	25
<i>A Priority Model for Named Entities</i> Lorraine Tanabe and W. John Wilbur .....	33
<i>Human Gene Name Normalization using Text Matching with Automatically Extracted Synonym Dictionaries</i> Haw-ren Fang, Kevin Murphy, Yang Jin, Jessica Kim and Peter White .....	41
<i>Integrating Co-occurrence Statistics with Information Extraction for Robust Retrieval of Protein Interactions from Medline</i> Razvan Bunescu, Raymond Mooney, Arun Ramani and Edward Marcotte .....	49
<i>BIOSMILE: Adapting Semantic Role Labeling for Biomedical Verbs</i> Richard Tzong-Han Tsai, Wen-Chi Chou, Yu-Chun Lin, Cheng-Lung Sung, Wei Ku, Ying-Shan Su, Ting-Yi Sung and Wen-Lian Hsu .....	57
<i>Generative Content Models for Structural Analysis of Medical Abstracts</i> Jimmy Lin, Damianos Karakos, Dina Demner-Fushman and Sanjeev Khudanpur .....	65
<i>Exploring Text and Image Features to Classify Images in Bioscience Literature</i> Barry Rafkind, Minsuk Lee, Shih-Fu Chang and Hong Yu .....	73
<i>Mining biomedical texts for disease-related pathways</i> Andrey Rzhetsky .....	81
<i>Postnominal Prepositional Phrase Attachment in Proteomics</i> Jonathan Schuman and Sabine Bergler .....	82

## Poster Papers

<i>BioKI:Enzymes - an adaptable system to locate low-frequency information in full-text proteomics articles</i> Sabine Bergler, Jonathan Schuman, Julien Dubuc and Alexandr Lebedev .....	91
<i>A Graph-Search Framework for GeneId Ranking</i> William Cohen .....	93
<i>Semi-supervised anaphora resolution in biomedical texts</i> Caroline Gasperin .....	96
<i>Using Dependency Parsing and Probabilistic Inference to Extract Relationships between Genes, Proteins and Malignancies Implicit Among Multiple Biomedical Research Abstracts</i> Ben Goertzel, Hugo Pinto, Ari Heljakka, Michael Ross, Cassio Pennachin and Izabela Goertzel ..	104
<i>Recognizing Nested Named Entities in GENIA corpus</i> Baohua Gu .....	112
<i>Biomedical Term Recognition with the Perceptron HMM Algorithm</i> Sittichai Jiampojarn, Grzegorz Kondrak and Colin Cherry .....	114
<i>Refactoring Corpora</i> Helen L. Johnson, William A. Baumgartner, Jr., Martin Krallinger, K. Bretonnel Cohen and Lawrence Hunter .....	116
<i>Rapid Adaptation of POS Tagging for Domain Specific Uses</i> John E. Miller, Michael Bloodgood, Manabu Torii and K. Vijay-Shanker .....	118
<i>Extracting Protein-Protein interactions using simple contextual features</i> Leif Arda Nielsen .....	120
<i>Identifying Experimental Techniques in Biomedical Literature</i> Meeta Oberoi, Craig A. Struble and Sonia L. Sugg .....	122
<i>A Pragmatic Approach to Summary Extraction in Clinical Trials</i> Graciela Rosembat and Laurel Graham .....	124
<i>The Difficulties of Taxonomic Name Extraction and a Solution</i> Guido Sautter and Klemens Böhm .....	126
<i>Summarizing Key Concepts using Citation Sentences</i> Ariel S. Schwartz and Marti Hearst .....	134
<i>Subdomain adaptation of a POS tagger with a small corpus</i> Yuka Tateisi, Yoshimasa Tsuruoka and Jun'ichi Tsujii .....	136
<i>Bootstrapping and Evaluating Named Entity Recognition in the Biomedical Domain</i> Andreas Vlachos and Caroline Gasperin .....	138