

Zhiyong Lu, Ph.D.

CONTACT INFORMATION DHHS/NIH/NLM/NCBI *Tel:* (301) 594-7089
8600 Rockville Pike *Fax:* (301) 480-2288
Building 38A, Room 6N609 *E-mail:* luzh@ncbi.nlm.nih.gov
Bethesda, MD 20894 <http://www.ncbi.nlm.nih.gov/CBBresearch/Lu>

RESEARCH INTERESTS Bioinformatics, Computational Biology
Text Mining (Named Entity Recognition; Biomedical Ontology; Information Retrieval)
Machine Learning, Data Mining

EMPLOYMENT Staff Scientist, National Center for Biotechnology Information (NCBI), 2007 -

EDUCATION Ph.D. in Bioinformatics, University of Colorado Health Sciences Center, 2007
PhD Thesis: Text Mining on GeneRIFs, advised by Lawrence Hunter

M.S. in Computer Science, University of Alberta, 2003
MS Thesis: Predicting Protein Sub-cellular Localization from Homologs Using Machine Learning Algorithms, supervised by Duane Szafron and Russell Greiner

B.S. in Computer Science, Nanjing University, 2001

HONORS AND AWARDS Best performance in protein-protein interaction task in BioCreative II global competition, 2007
Conference (PSB 2007) Travel Award, Int'l Society of Computational Biology (ISCB), 2007
Conference (ISMB 2006) Travel Award, Int'l Society of Computational Biology (ISCB), 2006
Conference (PSB 2006) Travel Award, Int'l Society of Computational Biology (ISCB), 2006
Conference (ISMB 2005) Travel Award, Int'l Society of Computational Biology (ISCB), 2005
Professional Development Award, University of Colorado HSC Student Senete, 2005
Conference (PSB 2005) Travel Award, Int'l Society of Computational Biology (ISCB), 2005
Conference (Rocky 03) Travel Award, Univ of Colorado Center of Computational Biology, 2003
Undergraduate (merit-based) Student Scholarship, Nanjing University, 1997 - 2001

PEER REVIEWED PUBLICATIONS **Lu Z** and Wilbur WJ: Improving accuracy for identifying related PubMed queries by an integrated approach, *Journal of Biomedical Informatics*, in press.

Lu Z, Kim W, Wilbur WJ: Evaluation of Query Expansion Using MeSH in PubMed, *Information Retrieval*, in press.

Lu Z, Kim W, Wilbur WJ: Evaluating Relevance Ranking Strategies for MEDLINE Retrieval, *JAMIA*, 2009. (*recognized as one of the 9 outstanding papers from a total of 457 papers submitted to the AMIA 2008 Annual Symposium, and invited for publication in JAMIA*).

Morgan AA, **Lu Z**, Wang X, Cohen AM, Fluck J, Ruch P, Divoli A, Fundel K, Leaman R, Hakenberg J, Sun C, Liu H, Torres R, Krauthammer M, Lau WW, Liu H, Hsu C, Schuemie M, Cohen KB, Hirschman L: Overview of BioCreative II Gene Normalization, *Genome Biology*, 2008, 9(Suppl 2):S2

Baumgartner WA, Jr, **Lu Z**, Johnson HL, Caporaso JG, Paquette J, Lindemann A, White EK, Medvedeva O, Cohen KB, Hunter L: Concept recognition for extracting protein interaction relations from biomedical text, *Genome Biology*, 2008, 9(Suppl 2):S8

Bethard S, **Lu Z**, Martin JH, Hunter L: Semantic Role Labeling for Protein Transport Predicates, *BMC Bioinformatics*, 9:277, 2008

Hunter L, **Lu Z**, Firby J, Baumgartner WA, Jr., Johnson HL, Ogren PV, Cohen KB: OpenDMAP: An open-source ontology-driven concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interaction, and cell-specific gene expression, *BMC Bioinformatics*, 9(1):78, 2008

Baumgartner W, **Lu Z**, Johnson HL, Caporaso JG, Paquette J, Lindemann A, White EK, Medvedeva O, Cohen KB, and Hunter L: An integrated approach to concept recognition in biomedical text, In *Proceedings of the Second BioCreative Challenge Evaluation Workshop*, 2007

Lu Z, Cohen KB, and Hunter L: GeneRIF Quality Assurance as Summary Revision, in *Proceedings of the Pacific Symposium on Biocomputing* 12:269-280, 2007

Caporaso JG, Baumgartner W, Kim H, **Lu Z**, Johnson H, Medvedeva O, Lindemann A, Fox L, White EK, Cohen KB, Hunter L: Concept Recognition, Information Retrieval, and Machine Learning in Genomics Question-Answering, in *Proceedings of The Fifteenth Text Retrieval Conference*, (TREC 2006)

Lu Z, Bada M, Ogren P, Cohen KB, and Hunter L: Improving Biomedical Corpus Annotation Guidelines, in *Proceedings of the Joint BioLink and 9th Bio-Ontologies Meeting*, 5 August 2006; Fortaleza, Brazil

Lu Z, Cohen KB, and Hunter L: Finding GeneRIFs via Gene Ontology Terms, in *Proceedings of the Pacific Symposium on Biocomputing*, 3-7 January 2006, Maui, Hawaii. 2006:52-63

Johnson HL, Cohen KB, Baumgartner WA, **Lu Z**, Bada M, Kester T, Kim H, and Hunter L: Evaluation of Lexical Methods for Detecting Relationships Between Concepts From Multiple Ontologies, in *Proceedings of the Pacific Symposium on Biocomputing*, 3-7 January 2006, Maui, Hawaii. 2006:28-49

Lu Z and Hunter, L: GO Molecular Function Terms are Predictive of Subcellular Localization, in *Proceedings of the Pacific Symposium on Biocomputing*, 4-8 January 2005, Big Island of Hawaii. 2005:151-161

Szafron D, Lu P, Greiner R, Wishart DS, Poulin B, Eisner R, **Lu Z**, Anvik J, Macdonell C, Fyshe A, Meeuwis D: Proteome Analyst: custom predictions with explanations in a web-based tool for high-throughput proteome annotations. *Nucleic Acids Research* 2004 Jul 1, 32:W365-71

Lu Z, Szafron D, Greiner R, Lu P, Wishart DS, Poulin B, Anvik J, Macdonell C, Eisner R: Predicting subcellular localization of proteins using machine-learned classifiers. *Bioinformatics*. 2004 Mar 1; 20(4): 547-56

Szafron D, Greiner R, Lu P, Wishart DS, **Lu Z**, Poulin B, Eisner R, Anvik J, Macdonell C: Proteome Analyst Transparent High-throughput Protein Function Prediction. In *Proceedings of the International Conference on Machine Learning, Workshop on Machine Learning in Bioinformatics*, August 2003, Washington DC, USA

Sander J, Qin X, **Lu Z**, Niu N and Kovarsky A: Automatic Extraction of Clusters from Hierarchical Clustering Representations, in *Proceedings of The 7th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, April 2003, Seoul Korea.

Lu Z, Xie N, Wilbur WJ: Are you reading the right journals, *submitted*.

- TUTORIALS **Lu Z** and Cohen KB: Introduction to biomedical natural language processing, *The Asia Pacific Bioinformatics Conference (APBC) 2009*, Beijing, China. January, 2009.
- REFEREE WORK
AND PROGRAM
COMMITTEE **Journals:**
Bioinformatics
BMC Bioinformatics
Journal of Biomedical Informatics
- Conferences:**
International Joint Conferences on Artificial Intelligence (IJCAI), Pasadena, CA, 2009
ISMB/ECCB 2009 Stockholm, Sweden, 2009
The Asia Pacific Bioinformatics Conference (APBC), 2009
NAACL-HLT BioNLP workshop, Boulder, CO 2009
ACL-HLT BioNLP workshop, Columbus, Ohio 2008
NAACL-HLT BioNLP workshop, Brooklyn, NY 2006
Pacific Symposium on Biocomputing (PSB), Maui, HI 2006
- INVITED LECTURES *Evaluating IR Techniques for Enhancing MEDLINE Retrieval*, Computational Bioscience Program, University of Colorado School of Medicine, December, 2008.
- Evaluating and Enhancing NCBI Textual Resources*, CBCB Seminar, University of Maryland College Park, November 2008
- Evaluating and Enhancing NCBI Textual Resources*, AI Seminar, Department of Computer Science, University of Alberta, April 2008
- Text Mining on GeneRIFs*, CLIP Colloquium, University of Maryland College Park, November 2007
- Text Mining for Biology*, Center for Bioinformatics, Peking University, June 2007
- Text Mining for Biology*, School of Life Sciences, Suzhou University, February 2007
- Extracting Protein Transport Information from GeneRIFs*, Bioinformatics supergroup, University of Colorado at Boulder, March 2006
- Case Study: Protein Translocation*, Biomedical Ontology Session Discussion Panel, Pacific Symposium on Biocomputing, January 2005
- OTHER SELECTED
PRESENTATIONS *Improving Biomedical Corpus Annotation Guidelines*, oral presentation at the Joint BioLink and 9th Bio-Ontologies Meeting, 2006; Fortaleza, Brazil
- Frame-based Analysis of Complex Relationships in Biomedical Text: Annotating and Extracting Protein Transport Information in GeneRIFs*, poster presentation at the 14th International Conference on Intelligent Systems for Molecular Biology, Fortaleza, Brazil, 2006
- Finding GeneRIFs via GO Annotations*, oral presentation at the Pacific Symposium on Biocomputing (PSB), Maui, Hawaii, 2006
- Mining Protein Transport Data from GeneRIFs*, poster presentation at the 3rd Annual Rocky Mountain Regional Bioinformatics Conference, Snowmass, Colorado, 2005
- PTKB: Protein Transport Knowledge Base A NLP Approach*, poster presentation at the 13th International Conference on Intelligent Systems for Molecular Biology, Detroit, Michigan, 2005

GO Molecular Function Terms are Predicative of Subcellular Localization, poster presentation at the Pacific Symposium on Biocomputing (PSB), Big island, Hawaii, 2005

Using Proteome Analyst for Protein Subcellular Localization Predication, poster presentation at the 3rd International Conference of the Canadian Proteomics Initiative, Vancouver, BC, 2003

Learning to Predict Protein Function from Sequence, poster presentation at the 10th International Conference on Intelligent System for Molecular Biology, Edmonton, Alberta, 2002

TEACHING
EXPERIENCE

05/06-11/06 University of Colorado School of Medicine Graduate Student Supervisor
Supervise summer intern students and rotation graduate students

08/04-11/04 University of Colorado School of Medicine Teaching Assistant
BIOI 7711: Introduction to Bioinformatics (graduate level class)

09/02-12/02 University of Alberta Teaching Assistant
CMPUT 366: Introduction to Artificial Intelligence

01/02-04/02 University of Alberta Teaching Assistant
CMPUT 201: Practical Programming Methodology

09/01-12/01 University of Alberta Teaching Assistant
CMPUT 101: Introduction to Computing

SERVICE AND
MEMBERSHIPS

Member of the Association of Computational Linguistics (ACL), 2008 -
Member of the International Society of Computational Biology (ISCB), 2005 -
Student Representative in Department Faculty Meeting, University of Colorado, 2004
Treasurer-elect, Computer Science Graduate Student Association, University of Alberta, 2003
Councilor-elect, Graduate Student Association, University of Alberta, 2003
Member of the American Association for Artificial Intelligence (AAAI), 2002