

- *Research Assistant*
 Computer Science Department, University of Minnesota, Twin Cities Sep. 2004 - Sep. 2009
 Computer Science Department, University of Minnesota, Duluth Sep. 2002 - Sep. 2004
 - Conducted natural language processing research in:
 - * The development of unsupervised and supervised machine learning methods for the task of disambiguating ambiguous words in biomedical text.
 - * The development of statistical methods for identifying multi-word expressions and statistically relevant terms in biomedical, clinical and general English text.
 - Tasks included hypothesis definition, experimental design and implementation, data analysis and publication writing.
- *Teaching Assistant*
 Computer Science Department, University of Minnesota, Twin Cities Sep. 2004 - May 2006
 Computer Science Department, University of Minnesota, Duluth Sep. 2002 - May 2004
 - Taught labs and recitation sessions for undergraduate computer science courses.
 - Developed course, lab and recitation material; including homework, quizzes and exams.
- *Student Research Scientist*
 National Library of Medicine, National Institute of Health, Bethesda, MD Mar. 2008 - Sep. 2008
 - Worked in a team to research and develop unsupervised methods to determine the sense of an ambiguous word in biomedical text for the purpose of improving the automatic indexing of biomedical and clinical documents.
- *Student Research Scientist*
 Thomson Legal and Regulatory, Eagan, MN May 2006 - Aug 2006
 May 2005 - Aug 2005
 - Worked in a team to research and develop supervised learning methods to extract job information from legal news briefs for the purpose of automatically updating legal databases.
- *Student Research Scientist*
 Mayo Clinic, Rochester, MN Jun. 2004 - Aug. 2004
 Jun. 2003 - Aug. 2003
 - Worked in a team which conducted natural language processing research in:
 - * The development of statistical methods to automatically correct misspelled and misused words in clinical records for the purpose of normalization.
 - * The development of statistical methods to automatically identify statistically relevant collocations and terms in clinical records for the purpose of terminology building.
 - * The development of statistical methods to automatically parse terms in clinical text.
- *Computer Programmer*
 Chemical Engineering Department, University of Minnesota, Duluth Jan. 2002 - May 2002
 - Designed and implemented Windows CE applications for curriculum aids for chemical engineering students.
- *Computer Programmer*
 Computer Science Department, University of Minnesota, Duluth Jun. 2001 - Jul. 2001
 - Worked in a team to research available solutions to provide search engine capabilities for the department, and subsequently integrate the search engine Lucene into the Computer Science web page.
- *Computer Programmer*
 Maurices, Inc., Duluth, MN May 2000 - Jun. 2001
 - Developed applications in C/C++ using PL/SQL to access an ORACLE Database to generate reports for marketing and financial analysis.
 - Assisted in the migration of Mainframe to IBM AIX Platform by converting existing COBOL programs into C/C++.

- *Undergraduate Work Study*
Large Lakes Observatory, University of Minnesota, Duluth Dec. 1997 - May 2000
 - Collected scientific conductivity, temperature and depth data aboard research vessels in Lake Superior.
 - Procured equipment and supplies required by the lab to conduct their experiments aboard the research vessels.
 - Designed, implemented and tested a software package written in Matlab to graphically display temperature, depth and conductivity data acquired from Lake Superior.
- *United States Air Force* Sep. 1993 - Sep. 1997
 - Stationed: Aviano Airbase, Italy (2.5 yr) and Osan Airbase, South Korea (1 yr).

Refereed Journal Publications

- Jimeno-Yepes, A. & McInnes, B.T. & Aronson, A. (2011) Exploiting MeSH Indexing in MEDLINE to Generate a Data Set for Word Sense Disambiguation. *BMC Bioinformatics*. 2011 Jun 2;12(1):223. [Epub ahead of print].
- Jimeno-Yepes, A. & McInnes, B.T. & Aronson, A. (2011) Collocation Analysis for UMLS Knowledge-based Word Sense Disambiguation. *BMC Bioinformatics* 2011, 12(Suppl 3):S4.
- Pakhomov, S. & Pedersen, T. & McInnes, B.T. & Melton, G.B. & Ruggieri, A. & Chute, C.G. (2011) Towards a Framework for Developing Semantic Relatedness Reference Standards. *Journal of Biomedical Informatics*. 44(2), pp. 251-265.

Refereed Conference Publications

- Liu, Y. & McInnes, B.T. & Pedersen, T. & Melton, G.B. & Pakhomov, S. (2012) Semantic Relatedness Study Using Second Order Co-occurrence Vectors Computed with Biomedical Corpora, UMLS and WordNet. To Appear in the *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. Miami, Florida.
- McInnes, B.T. & Pedersen, T. & Liu, Y. & Melton, G.B. & Pakhomov, S. (2011) Knowledge-based Method for Determining the Meaning of Ambiguous Biomedical Terms Using Information Content Measures of Similarity. To Appear in the *Proceedings of the Annual Symposium of the American Medical Informatics Association*. Washington, DC.
- Zhang, R. & Pakhomov, S. & McInnes, B.T. & Melton, G.B. (2011) Evaluating Measures of Redundancy in Clinical Texts. To Appear in the *Proceedings of the Annual Symposium of the American Medical Informatics Association*. Washington, DC.
- McInnes, B.T. & Pedersen, T. & Liu, Y. & Pakhomov, S. & Melton, G.B. (2011) Using Second-order Vectors in a Knowledge-based Method for Acronym Disambiguation. Appears in the *Proceedings of the Fifteenth Conference on Computational Natural Language Learning (CoNLL)*. Portland, Oregon.
- Pakhomov, S. & McInnes, B.T. & Adam, T. & Liu, Y. & Pedersen, T. & Melton, G.B. (2010) Semantic Similarity and Relatedness between Clinical Terms: An Experimental Study. Appears in the *Proceedings of the Annual Symposium of the American Medical Informatics Association*. Washington, DC.
- McInnes, B.T. & Pedersen, T. & Pakhomov, S. (2009) UMLS-Interface and UMLS-Similarity : Open Source Software for Measuring Paths and Semantic Similarity. Appears in the *Proceedings of the Annual Symposium of the American Medical Informatics Association*. San Francisco, CA.
- McInnes, B.T. & Pedersen, T. & Carlis, J. (2007) Using UMLS Concept Unique Identifiers (CUIs) for Word Sense Disambiguation in the Biomedical Domain. Appears in the *Proceedings of the Annual Symposium of the American Medical Informatics Association*. Chicago, IL.

Refereed Workshop Publications

- Pedersen, T. & Banerjee, S. & McInnes, B.T. & Kohli, S. & Joshi, M. & Liu, Y. (2011) The Ngram Statistics Package (Text::NSP) - A Flexible Tool for Identifying Ngrams, Collocations, and Word Associations. Appears in the *Proceedings of Multiword Expressions : from Parsing and generation to the Real World (MWE), an ACL HLT 2011 Workshop*. Portland, Oregon. (Demonstration System)

- Melton, G.B. & Moon, S. & McInnes, B.T. & Pakhomov, S. (2010) Automated Identification of Synonyms in Biomedical Acronym Sense Inventories. Appears in the *Proceedings of the Louhi Workshop at the North American Association of Computational Linguistics (NAACL)*. Los Angeles, CA.
- McInnes, B.T. (2008) An Unsupervised Vector Approach to Biomedical Term Disambiguation: Integrating UMLS and Medline. Appears in the *Proceedings of the Association for Computational Linguistics Student Research Workshop (ACL-SRW)*, Columbus, Ohio.
- McInnes, B.T. & Pedersen, T. & Pakhomov, S. (2007) Determining the Syntactic Structure of Medical Terms in Clinical Notes. Appears in the *Proceedings of the ACL Workshop BioNLP 2007: Biological, translational and clinical language processing*. Prague, Czech Republic.

Workshop Publications

- McInnes, B.T. (2008) Using CuiTools to Identify Obesity and its Co-morbidities in Discharge Summaries. Appears in the *Proceedings of the Second i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, Washington, DC.
- Schilder F. & McInnes, B.T. (2006) TLR at DUC 2006: Approximate Tree Similarity and a New Evaluation Regime. Appears in the *Proceedings of the Document Understanding Conference (DUC)*, New York, NY, USA.
- Schilder F. & McInnes, B.T. (2006) Word and Tree-based Similarities for Textual Entailment. Appears in the *Proceedings of the Second PASCAL Challenges Workshop on Recognizing Textual Entailment (RTE-2)*, Venice, Italy.
- Schilder, F. & McCullom, A. & Zhou, A. & McInnes, B.T. (2005) TLR at DUC: Tree Similarity. Appears in the *Proceedings of the Document Understanding Conference (DUC)*, Vancouver, Canada.
- McInnes, B.T. & Pedersen, T. (2003) The Duluth Word Alignment System. Appears in the *Proceedings of the NAACL Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*. Edmonton, Canada.

Published Abstracts

- Pakhomov S. & McInnes, B.T. (2005) Resolving Structural Ambiguity of Medical Terms with Statistical Model Fitting. Appears in the *Proceedings of the Linguistic Society of America (LSA)*. Oakland, CA. (panel presentation)
- McInnes, B.T. & Pakhomov, S. & Pedersen T. & Chute, C. (2004) Incorporating Bigram Statistics to Spelling Correction Tools. Appears in the *Proceedings of the 11th World Congress on Medical Informatics (MEDINFO)*. San Francisco, CA. IOS Press. (poster presentation)

Conference Reviewing

- International Health Informatics Symposium, January 2012, Miami. [6 papers]
- Association of Computational Linguistics (ACL) Conference, June 2011, Portland. [4 papers]
- Language Resources and Evaluation (LREC) Conference, May 2010, Malta. [3 papers]
- Association of Computational Linguistics (ACL) Workshop on BioNLP, June 2008, Boulder, CO. [1 paper]
- Association of Computational Linguistics (ACL) Workshop on Building and Using Parallel Texts: Data Driven MT and Beyond, June 2005, Ann Arbor, MI. [3 papers]

Participation in Shared Tasks and Comparative Evaluations

- Second i2b2 Shared-Task and Workshop Challenges in Natural Language Processing for Clinical Data Obesity Challenge (2008). Organized by the Informatics for Integrating Biology and the Bedside, i2b2.
- Medical NLP Challenge - Classifying Clinical Free Text Using Natural Language Processing (2007). Organized by the Computational Medicine Center.
- Second PASCAL Challenges Workshop on Recognizing Textual Entailment (2006). Organized by Pattern Analysis, Statistical Modeling and Computational Learning.

- DUC Shared Task: Automatic Summarising Evaluation Programme (2005 & 2006). Organized by the National Institute of Standards.
- ACL Shared Task: Building and Using Parallel Text: Data-driven Machine Translation and Beyond (2005). Organized by the Association for Computational Linguistics.

Selected Software Packages

- UMLS::Similarity (http://atlas.ahc.umn.edu/cgi-bin/umls_similarity.cgi) is a suite of freely available Perl modules that implement a number of semantic similarity measures in order to calculate the similarity between two biomedical or clinical concepts in the Unified Medical Language System.
- UMLS::SenseRelate (<http://search.cpan.org/dist/UMLS-SenseRelate>) is a freely available suite of Perl modules that performs knowledge-based word sense disambiguation.
- UMLS::Interface (<http://search.cpan.org/dist/UMLS-Interface>) is a freely available Perl interface to the Unified Medical Language System.
- CuiTools (<http://cuitools.sourceforge.net/>) is a freely available package of Perl programs that implements unsupervised and supervised methods for the task of word sense disambiguation, and classification of biomedical and clinical text.

Invited Talks

- Institute of Health Informatics, University of Minnesota, Twin Cities. March 2011.
- Thomson-Reuters, Eagan, MN. September 2009.
- National Library of Medicine, Bethesda, MD. September 2008.
- Mayo Clinic, Rochester, MN. August 2003 and August 2004.

Honors and Awards

- Graduate Assistance in Areas of National Need (GAANN) Fellowship, 2006-2009.
Awarded by the Computer Science and Engineering Department, University of Minnesota, Twin Cities.
- National Library of Medicine Research Participation Fellowship, 2008.
Awarded by the National Library of Medicine, National Institute of Health, Bethesda, MD.
- Most Outstanding Teaching Assistant , 2004.
Awarded by the Graduate School, University of Minnesota, Duluth.
- National Science Foundation scholarship, 2001-2002.
Awarded by the Computer Science Department, University of Minnesota, Duluth.

References available upon request.