Curriculum Vitae
Bridget T. McInnes

August 5, 2014

Education

- **University of Minnesota, Twin Cities**
  Ph.D., Computer Science, September 2009
  Dissertation: *Supervised and Knowledge-based Methods for Disambiguating Terms in Biomedical Text using the UMLS and MetaMap*

- **University of Minnesota, Duluth**
  M.S., Computer Science, December 2004
  Thesis: *Extending the Log-Likelihood Ratio to Improve Collocation Identification*

- **University of Minnesota, Duluth**
  B.S., Computer Science, May 2002

Research Interests

My research interests lie in the area Natural Language Processing and Computational Linguistics. My main research focus is in *semantics* with a particular emphasis on unsupervised knowledge-based algorithms. The main goal of my work is to have a serious impact on developing semantic tools that can be utilized for real-world problems involving the processing of natural language.

Journal Articles

[6] **McInnes, B.T.** & Pedersen, T. (2013) Evaluating Measures of Semantic Similarity and Relatedness to Disambiguate Terms in Biomedical Text. To Appear in the *Journal of Biomedical Informatics.*

[5] **McInnes, B.T.** & Stevenson, M. (2013) Determining the Difficulty of Word Sense Disambiguation. To Appear in the *Journal of Biomedical Informatics.*

[4] Pakhomov, S. & **McInnes, B.T.** & Lamba, J. & Liu, Y. & Melton, G.B. & Ghodke,Y. & Bhise, N. & Lamba, V. & Birnbaum, A.K. (2012) Using PharmGKB to Train Text Mining Approaches for Identifying Potential Gene Targets for Pharmacogenomic Studies. *Journal of Biomedical Informatics*, 45(5):862-9.

[3] 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*, 12(1):223.

[2] Jimeno-Yepes, A. & **McInnes, B.T.** & Aronson, A. (2011) Collocation Analysis for UMLS Knowledge-based Word Sense Disambiguation. *BMC Bioinformatics*, 12(Suppl 3):S4.

[1] Pakhomov, S. & Pedersen, T. & **McInnes, B.T.** & Melton, G. & Ruggieri, A. & Chute, C. (2011) Towards a Framework for Developing Semantic Relatedness Reference Standards, *Journal of Biomedical Informatics*, 44(2):251-65.

Refereed Conference Publications

10 **McInnes, B.T.** & Pedersen, T. & Liu, Y. & Melton, G.B. & Pakhomov, S. (2014) U-path: An undirected path-based measure of semantic similarity. To appear in the Proceedings of the Annual Symposium of the American Medical Informatics Association (AMIA), November 2014, Washington, D.C.

[9] Vogel, K. and **McInnes, B.T.** (2013) Understanding European Institutional Policy Discourse on the Council of Europe's Convention on Preventing and Combating Violence Against Women through Automated Topic-based Content Analysis. In *Proceedings of the Midwest Political Science Association Annual Conference*, April 2013, Chicago, IL.

[9] Bill, R.W. & Liu, Y. & **McInnes, B.T.** & Melton, G.B. & Pedersen, T. & and Pakhomov, S. (2012) Evaluating Semantic Relatedness and Similarity Measures with Standardized Med-DRA Queries. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*, November 2012, pp. 587 – 595, Chicago IL.

[8] Vogel, K. and **McInnes, B.T.** (2012) Understanding the Regional and Constitutional Impact of CEDAW Using Automated Content Analysis of Nation-state Constitutions. In *Proceedings of the Midwest Political Science Association Annual Conference*, April 2012, Chicago, IL.

[7] 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. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, January 2012, pp. 363–371, Miami, FL.

[6] **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. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*, October 2011, pp. 895–904, Washington, DC.

[5] Zhang, R. & Pakhomov, S. & **McInnes, B.T.** & Melton, G.B. (2011) Evaluating Measures of Redundancy in Clinical Texts. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*, October 2011, pp. 1612–1620, Washington, DC.

[4] **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. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning*, June 2011, pp. 145–153, Portland, OR.

[3] 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. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*. November 2010, pp. 572–576, Washington, DC.

[2] **McInnes, B.T.** & Pedersen, T. & Pakhomov, S. (2009) UMLS-Interface and UMLS-Similarity: Open Source Software for Measuring Paths and Semantic Similarity. In *Proceedings of the*

Annual Symposium of the American Medical Informatics Association. November 2009, pp. 431–435, San Francisco, CA.

[1] **McInnes, B.T.** & Pedersen, T. & Carlis, J. (2007) Using UMLS Concept Unique Identifiers (CUIs) for Word Sense Disambiguation in the Biomedical Domain. In *Proceedings of the Annual Symposium of the American Medical Informatics Association*, November 2007, pp. 533–537, Chicago, IL.

Refereed Workshop Publications

[3] Melton, G.B. & Moon, S. & **McInnes, B.T.** & Pakhomov, S. (2010) Automated Identification of Synonyms in Biomedical Acronym Sense Inventories. In *Proceedings of the Louhi Workshop at the North American Association of Computational Linguistics*, June 2010, pp. 46–52, Los Angeles, CA.

[2] **McInnes, B.T.** (2008) An Unsupervised Vector Approach to Biomedical Term Disambiguation: Integrating UMLS and Medline. In *Proceedings of the Association for Computational Linguistics Student Research Workshop*, June 2008, pp.49–54, Columbus, Ohio.

[1] **McInnes, B.T.** & Pedersen, T. & Pakhomov, S. (2007) Determining the Syntactic Structure of Medical Terms in Clinical Notes. In *Proceedings of the ACL Workshop BioNLP 2007: Biological, translational and clinical language processing*, June 2007, pp. 9–16, Prague, Czech Republic.

Demonstration Systems and Tutorials

[3] **McInnes, B.T.** & Pedersen, T. & Liu, Y. & Pakhomov, S. & Melton, G.B. (2013) UMLS::Similarity: Measuring the Relatedness and Similarity of Biomedical Concepts. In *Proceedings of the North American Association of Computational Linguistics Demonstration Systems*, June 2013, pp. 28–31, Atlanta, GA.

[2] Pedersen, & T. **McInnes, B.T.** & Pakhomov, S. & Liu, Y. (2012) Measuring the Similarity and Relatedness of Concepts in the Medical Domain : IHI 2012 Tutorial Overview. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, January 2012, pp.879, Miami, FL.

[1] 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.

Workshop Publications

[5] **McInnes, B.T.** (2008) Using CuiTools to Identify Obesity and its Co-morbidities in Discharge Summaries. In *Proceedings of the Second i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data*, Washington, DC.

[4] Schilder F. & **McInnes, B.T.** (2006) TLR at DUC 2006: Approximate Tree Similarity and a New Evaluation Regime. In *Proceedings of the Document Understanding Conference (DUC)*, New York, NY.

[3] Schilder F. & **McInnes, B.T.** (2006) Word and Tree-based Similarities for Textual Entailment. In *Proceedings of the Second PASCAL Challenges Workshop on Recognizing Textual Entailment (RTE-2)*, Venice, Italy.

[2] Schilder, F. & McCullom, A. & Zhou, A. & **McInnes, B.T.** (2005) TLR at DUC: Tree Similarity. In *Proceedings of the Document Understanding Conference (DUC)*, Vancouver, Canada.

[1] **McInnes, B.T.** & Pedersen, T. (2003) The Duluth Word Alignment System. In *Proceedings of the NAACL Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, Edmonton, Canada.

Published Abstracts

[2] Pakhomov S. & **McInnes, B.T.** (2005) Resolving Structural Ambiguity of Medical Terms with Statistical Model Fitting. In *Proceedings of the Linguistic Society of America (LSA)*, Oakland, CA.

[1] **McInnes, B.T.** & Pakhomov, S. & Pedersen T. & Chute, C. (2004) Incorporating Bigram Statistics to Spelling Correction Tools. In Proceedings of the 11th World Congress on Medical Informatics (MEDINFO) San Francisco, CA.

Participation in Shared Tasks and Comparative Evaluations

[5] 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.

[4] Medical NLP Challenge - Classifying Clinical Free Text Using Natural Language Processing (2007). Organized by the Computational Medicine Center.

[3] Second PASCAL Challenges Workshop on Recognizing Textual Entailment (2006). Organized by Pattern Analysis, Statistical Modeling and Computational Learning.

[2] DUC Shared Task: Automatic Summarising Evaluation Programme (2005 & 2006). Organized by the National Institute of Standards.

[1] ACL Shared Task: Building and Using Parallel Text: Data-driven Machine Translation and Beyond (2005). Organized by the Association for Computational Linguistics.

Invited Talks

[4] Institute of Health Informatics, University of Minnesota, Twin Cities. March 2011.

[3] Thomson-Reuters, Eagan, MN. September 2009.

[2] National Library of Medicine, Bethesda, MD. September 2008.

[1] Mayo Clinic, Rochester, MN. August 2003 and August 2004.

Honors and Awards

[**4**] Graduate Assistance in Areas of National Need (GAANN) Fellowship, 2006-2009. Awarded by the Computer Science and Engineering Department, University of Minnesota, Twin Cities.

[**3**] National Library of Medicine Research Participation Fellowship, 2008. Awarded by the National Library of Medicine, National Institute of Health, Bethesda, MD.

[**2**] Most Outstanding Teaching Assistant, 2004. Awarded by the Graduate School, University of Minnesota, Duluth.

[**1**] National Science Foundation scholarship, 2001-2002. Awarded by the Computer Science Department, University of Minnesota, Duluth.

Software Packages

[**4**] 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 quantify the similarity between two biomedical or clinical concepts in the Unified Medical Language System.

[**3**] UMLS::SenseRelate (http://search.cpan.org/dist/UMLS-SenseRelate) is a freely available suite of Perl modules that performs knowledge-based word sense disambiguation.

[**2**] UMLS::Interface (http://search.cpan.org/dist/UMLS-Interface) is a freely available Perl interface to the Unified Medical Language System.

[**1**] 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.

Conference Reviewing

[**2014**] (3) American Medical Informatics Association Symposium (AMIA);

[**2013**] (7) American Medical Informatics Association Symposium (AMIA); International Conference on Language Resources and Evaluation (LREC) Conference.

[**2012**] (11) International Health Informatics Symposium (IHI); American Medical Informatics Association Symposium (AMIA); IEEE International Workshop on Biomedical and Health Informatics.

[**2011**] (4) The 49th Annual Meeting of the Association of Computational Linguistics (ACL) Conference.

[**2010**] (3) International Conference on Language Resources and Evaluation (LREC) Conference.

[**2008**] (1) Association of Computational Linguistics (ACL) Workshop on BioNLP.

[**2005**] (3) Association of Computational Linguistics (ACL) Workshop on Building and Using Parallel Texts: Data Driven MT and Beyond.

Journal Reviewing

[3] Journal of Biomedical Informatics (2013, 1 article)

[2] Journal of Natural Language Engineering (2012, 1 article)

[1] Language and Linguistics Compass (2012, 1 article)