
KARIN M. VERSPOOR

Department of Computing and Information Systems
Level 8, Doug McDonnell Building
The University of Melbourne, VIC 3010 Australia

<http://textminingscience.com>
Mobile: +61 4 7840 8290
karin.verspoor@unimelb.edu.au

PROFESSIONAL EXPERIENCE

THE UNIVERSITY OF MELBOURNE – *Melbourne, Australia* **2014 – present**
Associate Professor, Department of Computing and Information Systems
Deputy Director, Health and Biomedical Informatics Centre

Teaching and Research in text and data mining.

- Focus on knowledge-based biomedical data analysis.
- Teaching of computer science subjects.
- Supervision of research students and post-doctoral fellows.

NATIONAL ICT AUSTRALIA – *Melbourne, Australia* **2011 – 2014**
Scientific Director, Health and Life Sciences 2012 – 2014
Biomedical Informatics team leader, Health and Life Sciences Business Area 2011 – 2014
Principal Researcher 2013 – 2014
Senior Researcher 2011 – 2013

Tasked with development of a biomedical informatics research program building on text mining capabilities.

- Research into knowledge-based biomedical data analysis.
- Engage biomedical industry and clinical partners to address challenging application areas.
- Collaborate with researchers in areas such as computational biology, data mining, and machine learning.
- Supervision of research staff.

UNIVERSITY OF COLORADO DENVER – *Denver, Colorado* **2008 – 2012**
Research Assistant Professor, Department of Pharmacology 2008 – 2011
Faculty, Computational Bioscience program 2008 – 2012

Tasked with development of a research program oriented to information extraction from the biomedical literature.

- Research into concept recognition and event extraction in the biomedical literature.
- Teach Computational Bioscience topics.
- Supervision of graduate students, postdoctoral researchers, and research staff.

LOS ALAMOS NATIONAL LABORATORY – *Los Alamos, New Mexico* **2003 – 2008**
Team Leader, Knowledge and Language Science team 2007 – 2008
Technical Staff Member, Knowledge and Information Systems Science team 2003 – 2007

Chosen to perform scientific research in computational methods for the extraction, representation, organization, synthesis, discovery and retrieval of knowledge in databases and information systems.

- Delivered information extraction systems for bioinformatics and homeland security applications, through integration of existing third-party tools and newly developed task-specific components.
- Identified and validated new unsupervised methods for word sense disambiguation in collaboration with a postdoctoral research assistant under my supervision.
- Streamlined parallel text processing efforts across four national laboratories for several homeland security applications by gaining adoption of a common architecture (UIMA) and minimizing redundancies.

APPLIED SEMANTICS, INC. – *Los Angeles, California* **2001 - 2002**
Senior Computational Linguist

Brought on to lead core technology development of an enterprise suite offering semantic meta-tagging, summarization and categorization of documents.

- Achieved improvements in system quality via redesign of existing word sense disambiguation and "page sensing" algorithms to incorporate NLP techniques such as reference resolution and shallow parsing.
- Augmented lexicon and ontology specificity through implementation of statistical techniques for semantic clustering and automatic acquisition of concept-contexts and selectional preferences.

KARIN M. VERSPOOR

WEBMIND, INC. – *New York, New York*

1998 - 2001

Director, Natural Language Engineering

Hired to design and implement the natural language module in the Webmind Artificial Intelligence system, a module aimed at deep semantic understanding of texts and intelligent, conversational query answering.

- Directed a team of nine scientists and programmers, leading conceptual and implementation-level design.
- Delivered a natural language processing module capable of fine-grained semantic representation of text, through integration of techniques derived from information retrieval algorithms, statistical parsing, unification-based analysis and information extraction.

MICROSOFT RESEARCH INSTITUTE, MACQUARIE UNIVERSITY – *Sydney, Australia*

1997 - 1998

Research Fellow: Conducted research in multilingual natural language generation, and served as Project Coordinator of the DDD (Dynamic Document Delivery) project, generating multilingual (Spanish, Chinese) textual descriptions of museum objects from an abstract database at varying levels of detail.

EDUCATION

UNIVERSITY OF MELBOURNE – *Melbourne, Australia*

expected 2016

Graduate Certificate in University Teaching

UNIVERSITY OF EDINBURGH – *Edinburgh, Scotland, UK*

1997

PhD, Cognitive Science and Natural Language

Thesis: “Contextually-Dependent Lexical Semantics”

UNIVERSITY OF EDINBURGH – *Edinburgh, Scotland, UK*

1994

MSc, Cognitive Science and Natural Language

Thesis: “A Cognitively-Relevant Lexical Semantics”

Awarded with Distinction

RICE UNIVERSITY – *Houston, TX*

1993

BA, Computer Science and Cognitive Sciences

Thesis: “What are the Characteristics of Emotional Metaphors?”

Graduated *Summa Cum Laude*

SCIENTIFIC AND TECHNICAL EXPERTISE

Biomedical informatics research experience:

- Concept and event extraction from the biomedical literature
- Machine learning and data mining for clinical data analysis
- Text mining for genetic variant interpretation
- Protein function prediction
- Formal methods in biological reasoning using bio-ontologies

Computational linguistics research experience:

- Information extraction, ontology-based methods for semantic analysis, lexical semantics, word sense disambiguation, coreference resolution and natural language generation.
- Language modeling, latent semantic analysis and terminology induction.
- Extensive background in theoretical linguistics and formal semantics.
- Application of machine learning methods

Extensive software engineering and data management experience, including:

- Programming languages: Java, Perl, C++, Prolog, UNIX scripting tools, some Python, some Lisp.
- Database systems: Using ORACLE, DB2, SQL Server for data storage and analysis. SPARQL and RDF.

INVITED PRESENTATIONS AND HONORS

Keynote Presentations

- Invited Keynote, Yonsei University workshop on Mining Big Text, Seoul, Korea, 02/2015. Drawing on Millions of Biomedical Journal Publications to do Predictive Biology.
- Invited Keynote Paper: Burnham D, Estival D, Cassidy S, Sefton P, [Verspoor, K.](#) Two Platforms for Research in Human Communication Science: The AusTalk Corpus and the Alveo Virtual Laboratory. Oriental Chapter of

KARIN M. VERSPOOR

COCOSDA (International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques) Thailand, 09/2014.

- Keynote speaker at the Medical Information Retrieval (MedIR) Workshop at SIGIR2014, “Practice-based Evidence in Medicine: Where Information Retrieval Meets Data Mining”, 07/2014.
- Yahoo! Distinguished Lecturer series, Columbia University Center for Computational Learning Systems “Semantics, Text, and Biomedical Knowledge Discovery: Challenges and Opportunities”, 02/2010.
- Keynote speaker at the Rocky Mountain Bioinformatics Conference, “Integrating Semantic Data Sources for Disease-implicated Gene Discovery”, 12/2007.
- Keynote speaker at the Recent Advances in Natural Language Processing conference, “Semantics, Text, and the cure for Cancer”, 09/2007, (also presented at MITRE Corp, 08/2007)

Honors and Invited Participation

- Invited participant, BioHackathon 2014, Miyagi, Japan, 11/2014.
- Invited participant, Google Faculty workshop on Language Understanding and Knowledge Discovery, Nanjing, China, 2013.
- Invited participant, BioHackathon 2013, Tokyo, Japan, 06/2013.
- Senior Fellow, Computing and Information Systems, The University of Melbourne, 2012-2014
- Adjunct Professor, Communication Studies Department, New Mexico State University, 2005-2006

Invited Seminars

- “A Birds' Eye View of Research in Biomedical Text and Data Mining at the University of Melbourne”, Invited presentation, Yonsei University workshop on Mining Big Text, Seoul, Korea, 02/2015.
- “Drawing on Millions of Biomedical Journal Publications to do Predictive Biology”, Invited presentation, 2015 International Conference on Big Data and Smart Computing (BigComp 2015), Jeju Island, Korea, 02/2015.
- “Drawing on Millions of Biomedical Journal Publications to do Predictive Biology”, Invited seminar, University of Melbourne Department of Mathematics and Statistics Big Data reading group, 10/2014.
- “Bioinformatics and data analytics for next-generation cancer care”, Invited speaker, The 5th Annual Integrated Cancer Centres Symposium, Melbourne, Australia, 12/2013.
- “A Web Service Annotation Framework for CTD Using the UIMA Concept Mapper”. BioCreative IV Workshop, Bethesda, MD, 10/2013.
- “Exploring the recovery of curated genetic variants through text mining”. Invited presentation to the National Center for Biotechnology Information, US National Institutes of Health and at the University of Maryland Baltimore County, 09/2013.
- “An assessment of the recovery of curated genetic variants through text mining”, Late Breaking Research presentation, Intelligent Systems for Molecular Biology meeting, Berlin, Germany, 07/2013.
- “Towards Standardization of Annotation Representation for Text and Other Web Resources”. Invited presentation at the Public Symposium of the BioHackathon 2013, Tokyo, Japan, 06/2013.
- “From Structured Data to Unstructured Text and Back: Experiments in Text Mining for Biomedicine”. Invited talks at The University of Melbourne (Computing and Information Systems seminar, 03/2013) and RMIT (Computer Science seminar, 03/2013)
- “Unlocking Biomedical Knowledge through Text Mining”, CS4HS conference, 11/2012
- “Text Mining improves Prediction of Protein Functional Sites”, Highlights track presentation, Intelligent Systems for Molecular Biology meeting, Long Beach, CA, 07/2012.
- “Literature-Enhanced Strategies for Biological Data Interpretation”, Baker IDI Heart & Diabetes Institute, 05/2012
- “Roles for Text Mining in Understanding Genetic Variation”, Human Variome Project seminar series, University of Melbourne, 05/2012
- “Beyond Search: Enabling Biomedical Knowledge Discovery through Natural Language Processing”. Invited talks at: Colorado State University (Computer Science Colloquium, 10/2011), Los Alamos National Laboratory (Center for Nonlinear Studies Seminar, 09/2011), NICTA Victoria Research Laboratory (04/2011), the University of the Basque Country (03/2011), Macquarie University Computer Science Department (02/2012)
- “Annotating the Biomedical Literature through Natural Language Processing”, invited workshop presentation, Open Annotation Collaboration phase II workshop.
- Panelist, “Health Informatics – Making a Difference Through Technology” session at the Grace Hopper Celebration of Women In Computing, Atlanta, GA, 09/2010.
- “Semantics, Text, and Biomedical Knowledge Discovery: Challenges and Opportunities”, invited presentation at the University of Cambridge (UK) Computing Laboratory, 09/2010
- “Information integration for biomedical knowledge discovery”, invited workshop presentation UCLA IPAM Search Engines reunion conference, Lake Arrowhead, CA, 06/2009.

KARIN M. VERSPOOR

- “Search Engines program summary”, invited workshop presentation UCLA IPAM Search Engines reunion conference, Lake Arrowhead, CA, 06/2009.
- “Statistics, Semantics, and Structure: examples of hybrid approaches to text processing problems”, invited presentations at U. Colorado Denver, Pacific Northwest National Laboratory and Thomson Research, Fall 2007
- “Exploiting contextual similarity for semantic prediction in the Biology domain”, invited workshop presentation, UCLA IPAM workshop on “Dynamic Searches and Knowledge Building”, 10/2007
- “Mining Knowledge to Build Social Networks”, invited workshop presentation UCLA IPAM “Social Data Mining and Knowledge Building” workshop, 11/2007
- “A Distributed Architecture for Artificial Intelligence”. Presented at the Association of Women in Science Central New Jersey Chapter meeting, 06/2000.
- “The art of being a successful technical manager”. Presented at the Association for Women In Computing New York Chapter meeting, 10/1999.
- “A Lexical Semantic Account of Prepositional Phrase Modification”. Presented at the 20th annual meeting of the DGfS (German Linguistics Society); Workshop on the Syntax of Adverbials, 02/1998.
- “Phrasal constructs, the lexicon, and multilingual generation”. Presented at the Hong Kong University of Science and Technology Computer Science department, 02/1998. Also presented at Logos Corporation (Mount Arlington, NJ) and the Columbia University Computer Science department.

TEACHING AND SUPERVISION

- Co-supervision of two PhD students at UC Denver (Christopher Funk, Benjamin Garcia); eight PhD students at The University of Melbourne (Qingyu Chen, Miji Choi, Long Duong, Nusrath Hameed, Patrick Pang, Simone Romano, Ping Ping Tan, Doris Hoogeveen)
- COMP30018/90049 Knowledge Technologies, The University of Melbourne (co-teaching (50%) SM2 2013, subject coordinator SM2 2014)
- Guest lecture in COMP10001 Foundations of Computing, The University of Melbourne (09/2013)
- Guest lectures on The Electronic Health Record and the need for Text Mining, ISYS90069 eHealth and Biomedical Informatics Systems, The University of Melbourne (04/2013 and 07/2014)
- Guest lectures on Biomedical Text Mining, INFO90001 eHealth and Biomedical Informatics Methods, The University of Melbourne (04/2013 and 11/2013)
- Australasian Language Technology Association Workshop, Invited Tutorial (with David Martinez and Hanna Suominen): “Biomedical Natural Language Processing”, 12/2012
- Supervision of post-doctoral work by Komandur Ravikumar and Haibin Liu (UC Denver, 2010-2012)
- Computational Bioscience PhD program, University of Colorado Denver, 2009-2011
 - Fall 2009, Fall 2010, Fall 2011 Computational Bioscience 7711: Sequence search and Alignment, Text mining & Biomedical Natural Language Processing, Preparing research manuscripts, How to present research orally
 - Spring 2010, 2011 Computational Bioscience 7712: special topics (Research in the Verspoor Lab)
 - Spring 2010, 2011 PHSC/BMST 7454, Structural Analysis of Biomolecules II; Fall 2010, 2011, PHCL 7600: Frontiers in Pharmacology lecture “Computational Prediction of Protein Structure”
- Linguistic Institute 2011, U. Colorado Boulder, guest lecturer, Computational Lexical Semantics course.
- Linguistics, University of Colorado Boulder, 2010
 - Spring 2010, Ling7800: Advanced Computational Linguistics: Lexical Semantics, guest lecturer. “Verbs for Bioinformatics”, “Statistical Methods for doing semantics”
- Recent Advances in Natural Language Processing, Invited tutorial (with Kevin B. Cohen): “Natural Language Processing and the Biomedical Domain”, 09/2007
- Teaching and tutoring in computer science and computational linguistics 1990-1998
 - Co-teaching, Introduction to Language Technology, Microsoft Research Institute, Sydney, 1998
 - Tutor/Demonstrator, MSc program, University of Edinburgh, UK, Centre for Cognitive Science
 - Fall 1995, Fall 1996: Head tutor, Logic Programming and Computational Linguistics courses.
 - Spring 1996: Head tutor, Computational Semantics course.
 - Fall 1994: Tutor, Logic Programming and Computational Linguistics courses.
 - Tutor/Marker, MSc program, University of Edinburgh, UK, Dept. of Artificial Intelligence, 1994-1995
 - Tutor and Marker, Introduction to Natural Language Processing course.
 - Teaching Assistant, undergraduate courses, Rice University, Dept. of Computer Science, 1990-1993
 - Spring 1993: Artificial Intelligence; Fall 1990, Spring 1992: Data Structures & C++ programming

KARIN M. VERSPOOR

JOURNAL PUBLICATIONS

1. Funk CS, Kahanda I, Ben-Hur A, Verspoor KM. (accepted 06-Feb-2015, to appear). Evaluating a variety of text-mined features for automatic protein function prediction with GOstruct. *BMC Bioinformatics*, special BioOntologies SIG supplement.
2. Groza T, Verspoor K. (accepted 11-Jan-2015, to appear) Assessing the impact of case sensitivity and term information gain on biomedical concept recognition. *PLoS ONE*.
3. Pang P, Verspoor K, Chang S, Pearce J. (2015). Conceptualising Health Information Seeking Behaviours and Exploratory Search: Result of A Qualitative Study. *Health and Technology*. doi:10.1007/s12553-015-0096-0.
4. Liu H, Verspoor K, Comeau DC, MacKinlay AD, Wilbur WJ. (accepted 07-Oct-2014, to appear) Optimizing graph-based patterns to extract biomedical events from the literature. *BMC Bioinformatics* special issue on BioNLP Shared Task 2013.
5. Krallinger M, Rabal O, Leitner F, Vazquez M, Salgado D, Lu Z, Leaman R, Lu Y, Ji D, Lowe DM, Sayle RA, Batista-Navarro RT, Rak R, Huber T, Rocktäschel T, Matos S, Campos D, Tang B, Xu H, Munkhdalai T, Ryu KH, Ramanan SV, Nathan S, Zitnik S, Bajec M, Weber L, Irmer M, Akhondi SA, Kors JA, Xu S, An X, Sikdar UK, Ekbal A, Yoshioka M, Dieb TM, Choi M, Verspoor K, Khabsa M, Giles CL, Liu H, Ravikummar KE, Lamurias A, Couto FM, Dai HJ, Tsai RT, Ata C, Can T, Usie A, Cruz J, Segura-Bedmar I, Martinez P, Oyarzabal J, Valencia A. The ChEMDNER corpus of chemicals and drugs and its annotation principles. *Journal of Cheminformatics* (2015) 7:S1, p. S2.
6. MacIntyre GJ, Jimeno Yepes A, Ong CS, Verspoor K*. (2014) Associating disease-related genetic variants in intergenic regions to the genes they impact. *PeerJ* 2:e639. <http://dx.doi.org/10.7717/peerj.639>
7. Martin Sanchez FJ, Verspoor K. Big Data in Medicine is driving Big Changes (2014) *International Medical Informatics Association, Yearbook in Medical Informatics 2014*, 9(1):14-20. doi:10.15265/IY-2014-0020.
8. Verspoor K*, Shatkay H, Hirschman L, Blaschke C, Valencia A. (2014) Summary of the BioLINK SIG 2013 meeting at ISMB/ECCB 2013. *Bioinformatics* 2014; doi:10.1093/bioinformatics/btu412
9. Comeau DC, Batista-Navarro RT, Dai HJ, Doğan RI, Jimeno A, Khare R, Lu Z, Marques H, Mattingly CJ, Neves M, Peng Y, Rak R, Rinaldi F, Tsai RTH, Verspoor K, Wieggers TC, Wu CH, Wilbur WJ. (2014) BioC Interoperability Track Overview. *Database, the Journal of Biological Databases and Curation*. 2014: bau053
10. Park DJ, Nguyen-Dumont T, Kang S, Verspoor K, Pope B. (2014) Annokey: an annotation tool based on key term search of the NCBI Entrez Gene database. *Source Code for Biology and Medicine*, 9:15. (doi:10.1186/1751-0473-9-15)
11. Funk C, Baumgartner WA, Garcia B, Roeder C, Bada M, Cohen KB, Hunter LE, Verspoor K*. (2014) Large-scale biomedical concept recognition: An evaluation of current automatic annotators and their parameters. *BMC Bioinformatics* 15(1):59.
12. Jimeno Yepes A, Verspoor K*. (2014) Mutation extraction tools can be combined for robust recognition of genetic variants in the literature. *F1000Research* 2014, 3:18 (doi: 10.12688/f1000research.3-18.v2)
13. Jimeno Yepes A, Verspoor K*. (2014) Literature mining of genetic variants for curation: Quantifying the importance of supplementary material. *Database: The Journal of Biological Databases and Curation*, bau003. doi:10.1093/database/bau003
14. Livingston K, Bada M, Hunter LE, Verspoor KM*. (2013) Representing Annotation Compositionality and Provenance for the Semantic Web. *Journal of Biomedical Semantics*, 4:38. doi:10.1186/2041-1480-4-38.
15. Comeau DC, Doğan RI, Ciccarese P, Cohen KB, Krallinger M, Leitner F, Lu Z, Peng Y, Rinaldi F, Torii M, Valencia A, Verspoor K, Wieggers TC, Wu CH, and Wilbur WJ. (2013) BioC: A Minimalist Approach to Interoperability for Biomedical Text Processing. *Database: The Journal of Biological Databases and Curation*, bat064. doi:10.1093/database/bat064
16. Liu H, Hunter L, Keselj V, Verspoor K*. (2013) Approximate Subgraph Matching-based Literature Mining for Biomedical Events and Relations. *PLoS ONE* 8(4):e60954.
17. Verspoor K*, Jimeno Yepes A, Cavedon L, McIntosh T, Herten-Crabb A, Thomas Z, Plazzer JP. (2013) Annotating the Biomedical Literature for the Human Variome. *Database: The Journal of Biological Databases and Curation*, bat019.
18. Radivojac P, Clark W, Oron TR, Schnoes AM, Wittkop T, Sokolov A, Graim K, Funk C, Verspoor K et al. (2013) A large-scale evaluation of computational protein function prediction. *Nature Methods*. doi: 10.1038/nmeth.2340
19. Sokolov A, Funk C, Graim K, Verspoor K, Ben-Hur A (2013) Combining Heterogeneous Data Sources for Accurate Functional Annotation of Proteins. *BMC Bioinformatics* 14(S3):S10, special issue on CAFA.
20. Lippincott T, Rimell L, Verspoor K; Korhonen A. (2013) Approaches to Verb Subcategorization for Biomedicine, *Journal of Biomedical Informatics* 46(2):212-227.

KARIN M. VERSPOOR

21. Rimell L, Lippincott T, Verspoor K; Johnson H, Korhonen A. (2013) Acquisition and Evaluation of Verb Subcategorization Resources for Biomedicine, *Journal of Biomedical Informatics* 46(2):228-237.
22. Verspoor K*, Cohen KB, Lanfranchi A, Warner C, Johnson HL, Roeder C, Choi JD, Funk C, Malenkiy Y, Eckert M, Xue N, Baumgartner Jr. WA, Bada M, Palmer M, Hunter LE (2012) A corpus of full-text journal articles is a robust evaluation tool for revealing differences in performance of biomedical natural language processing tools. *BMC Bioinformatics*. 13(207).
23. Ravikumar K, Liu H, Cohn JD, Wall ME, Verspoor K*. Literature mining of protein-residue associations with graph rules learned through distant supervision (2012) *Journal of Biomedical Semantics*. 3(S3):S2.
24. Bada, M, Eckert M, Evans D, Garcia K, Shipley K, Sitnikov D, Baumgartner WA, Cohen KB, Verspoor K, Blake JA, Hunter LE (2012). Concept Annotation in the CRAFT corpus. *BMC Bioinformatics*. 13(161).
25. Liu H, Christiansen T, Baumgartner Jr WA, and Verspoor K* (2012) BioLemmatizer: a lemmatization tool for morphological processing of biomedical text. *Journal of Biomedical Semantics*, 3:3.
26. Verspoor KM*, Cohn JD, Ravikumar KE, Wall ME* (2012) Text Mining Improves Prediction of Protein Functional Sites. *PLoS ONE* 7(2): e32171. (selected for Highlights Presentation at ISMB 2012).
27. Lu Z, Kao HY, Wei C, Huang M, Liu J, Kuo CJ, Hsu CN, Tsai RT, Dai HJ, Okazaki N, Cho H, Gerner M, Solt I, Agarwal S, Liu F, Vishnyakova D, Ruch P, Romacker M, Rinaldi F, Bhattacharya S, Srinivasan P, Liu H, Torii M, Matos S, Campos D, Verspoor K, Livingston KM, Wilbur WJ (2011) The gene normalization task in BioCreative III. *BMC Bioinformatics*, 12(Suppl 8):S2.
28. Kano Y, Bjerne J, Ginter F, Salakoski T, Buyko E, Hahn U, Cohen KB, Verspoor K, Roeder C, Hunter LE, Kilicoglu H, Bergler S, Van Landeghem S, Van Parys T, Van de Peer Y, Miwa M, Ananiadou S, Neves M, Pascual-Montano A, Ozgur A, Radev DR, Riedel S, Saetre R, Chun HW, Kim JD, Pyysalo S, Ohta T, and Tsujii J. (2011) U-Compare bio-event meta-service: compatible BioNLP event extraction services. *BMC Bioinformatics* 12:481.
29. Cohen KB*, Verspoor K*, Johnson H, Roeder C, Ogren P, Baumgartner W, White E, Tipney H, and Hunter L. (2011) High-precision biological event extraction: Effects of system and data. *Computational Intelligence*, 27(4), p. 681-701.
30. Cohen KB, Johnson HL, Verspoor K, Roeder C, Hunter LE. (2010) The structural and content aspects of abstracts versus bodies of full text journal articles are different. *BMC Bioinformatics*, 11:492.
31. Verspoor K*, Roeder C, Johnson H, Cohen KB, Baumgartner WA, Hunter L. (2010) Exploring species-based strategies for gene normalization. *Transactions on Computational Biology and Bioinformatics*. Jul-Sep; 7(3): 462-471. doi: 10.1109/TCBB.2010.48.
32. Roeder C, Jonquet C, Shah NH, Baumgartner WA, Verspoor K and Hunter L. (2010). A UIMA wrapper for the NCBO annotator. *Bioinformatics*; 26(14): 1800-1801. doi: 10.1093/bioinformatics/btq250
33. Verspoor K*, Cohen KB, Hunter L. (2009) The textual characteristics of traditional and Open Access scientific journals are similar. *BMC Bioinformatics* 10:183.
34. Verspoor K*, Dvorkin D, Cohen KB, Hunter L. (2009) Ontology quality assurance through analysis of term transformations. *Bioinformatics* 25(12):i77-i84; doi:10.1093/bioinformatics/btp195.
35. Abi-Haidar A, Kaur J, Maguitman A, Radivojac P, Retchsteiner A, Verspoor K, Wang Z, Rocha LM. (2008). Uncovering protein interaction in abstracts and text using a novel linear model and word proximity networks. *Genome Biology*, 9(Suppl 2):S11.
36. Verspoor K*, Cohn J, Mniszewski S, and Joslyn C. (2006). A Categorization Approach to Automated Ontological Function Annotation. *Protein Science*, v.15, pp.1544-1549.
37. Verspoor K*. (2005). Towards a semantic lexicon for biological language processing. *Comparative and Functional Genomics*, vol. 6, issue 1-2, p. 61-66. DOI: 10.1002/cfg.451.
38. Verspoor K*, Cohn J, Joslyn C, Mniszewski S, Rechtsteiner A, Rocha LM, Simas T. (2005). Protein Annotation as Term Categorization in the Gene Ontology using Word Proximity Networks. *BMC Bioinformatics* 2005, vol. 6 (Suppl 1).

MEDIA ARTICLES

39. Verspoor K. (2014) "Diving deep into data to crack the gene code on disease". *The Conversation*. 14 February 2014. <http://tiny.cc/0luzdx>
40. Verspoor K. (2014) "C'mon girls, let's program a better tech industry". *The Conversation*. 29 July 2014. <http://tiny.cc/1u9tkx>

KARIN M. VERSPOOR

TECHNICAL STANDARDS WORK

41. Ferrucci D, Lally A, Verspoor K, Nyberg E. (editors) Unstructured Information Management Architecture (UIMA) Version 1.0, OASIS Standard, 2 March 2009.

PROCEEDINGS, BOOKS AND BOOK CHAPTERS

42. Holzinger A, Schantl J, Schroettner M, Seifert C, Verspoor K. (2014) Biomedical Text Mining: Open Problems and Future Challenges. In Lecture Notes in Computer Science, vol. 8401, *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*, pp. pp 271-300.
43. Verspoor K. (2014) Roles for Text Mining in Protein Function Prediction. In Kumar VD and Tipney HJ (Eds.) *Methods in Molecular Biology: Biomedical Literature Mining*. Springer 1159:95-108.
44. Barbuto K, Schaper L, Verspoor K, editors. (2014) Proceedings of the Abstracts of the Scientific Stream at Big Data 2014. CEUR workshop proceedings, Vol. 1149. urn:nbn:de:0074-1149-2
45. Karimi S., Verspoor K. (Editors) (2013) Proceedings of the Australasian Language Technology Association Workshop 2013 (ALTA 2013). Brisbane, Australia. <http://www.aclweb.org/anthology/U13-1>
46. Cohen KB, Verspoor K. Applied Text Mining. (2013) In Dubitzky W, Wolkenhauer O, Yokota H, Cho KH (Eds.) *Encyclopedia of Systems Biology*.
47. Verspoor K, Cohen KB. Natural Language Processing. (2013) In Dubitzky W, Wolkenhauer O, Yokota H, Cho KH (Eds.) *Encyclopedia of Systems Biology*.
48. Verspoor K, Baumgartner WA. UIMA: The Unstructured Information Management Architecture (2013) In Dubitzky W, Wolkenhauer O, Yokota H, Cho KH (Eds.) *Encyclopedia of Systems Biology*.
49. Gessler D, Joslyn C, and Verspoor K. (2013) A posteriori Ontology Engineering for Data-Driven Science. In Critchlow, T. and Kleese-Van Dam, K. (Eds.), *Data Intensive Science*. Boca Raton, FL: Taylor and Francis.
50. Verspoor K, Cohen KB, Goertzel B, Mani I. (Editors) (2006). *Linking Natural Language Processing and Biology: Towards Deeper Biological Literature Analysis*. Proceedings of the HLT-NAACL BioNLP Workshop (BioNLP'06). New York, NY.

PEER-REVIEWED CONFERENCE PUBLICATIONS

51. Duong L, Cohn T, Verspoor K, Bird S, Cook P. (2014) What We Can Get From 1k Tokens? Case Study of Multilingual POS Tagging For Resource-poor Languages. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), pages 886–897. Doha, Qatar.
52. Burnham D, Estival D, Cassidy S, Sefton P, Verspoor K. (2014) Two Platforms for Research in Human Communication Science: The AusTalk Corpus and the Alveo Virtual Laboratory. Oriental Chapter of COCODA (International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques) 2014.
53. Romano S, Bailey J, Vinh NX, Verspoor K. (2014) Standardized Mutual Information for Clustering Comparisons: Why Normalization Should Not Be The Standard. The 31st International Conference on Machine Learning (ICML 2014), Beijing, China, June 2014.
54. Ofoghi B, Lopez Campos GH, Martin Sanchez FJ, Verspoor K*. (2014) Mapping biomedical vocabularies: A semi-automated term matching approach. International Conference on Informatics, Management and Technology in Healthcare 2014 (ICIMTH 2014).
55. Pang C, Chang S, Pearce J, Verspoor K. (2014) Online Health Information Seeking Behaviour: Understanding Different Search Approaches. The 18th Pacific Asia Conference on Information Systems (PACIS 2014), Chengdu, China, June 2014.
56. Ofoghi B, Lopez Campos GH, Verspoor K, Martin Sanchez FJ. (2014) BiomRKRS: A Biomarker Retrieval and Knowledge Reasoning System. Health Information and Knowledge Management conference (HIKM 2013), Auckland, NZ.
57. Verspoor K*, MacKinlay A, Cohn JA, Wall MA (2013) Detection of protein catalytic sites in the biomedical literature. Pacific Symposium on Biocomputing, Hawaii, USA. January 2013.
58. Liu H, Keselj V, Blouin C, Verspoor K*. (2012) Subgraph Matching-based Literature Mining for Biomedical Relations and Events, Nov 2-4, 2012. AAAI 2012 Fall Symposium on Information Retrieval and Knowledge Discovery in Biomedical Text. Arlington, VA, USA.
59. Ravikumar KE, Liu H, Cohn JD, Wall ME and Verspoor K* (2011). Pattern Learning through Distant Supervision for Extraction of Protein-Residue Associations in the Biomedical Literature. International Conference on Machine Learning and Applications, Honolulu, Hawaii.

KARIN M. VERSPOOR

60. Livingston K, Johnson HL, Verspoor K, Hunter LE. (2010) Leveraging Gene Ontology Annotations to Improve a Memory-Based Language Understanding System. Fourth IEEE International Conference on Semantic Computing (IEEE ICSC2010), p. 40-45.
61. Görg C, Tipney H, Verspoor K, Baumgartner WA, Cohen KB, Stasko J, and Hunter LE. (2010) Visualization and Language Processing for Supporting Analysis across the Biomedical Literature. In proceedings, 14th Intl Conference on Knowledge-Based and Intelligent Information & Engineering Systems, KES2010, Workshop on 3D visualization of natural language.
62. Cohen KB, Roeder C, Baumgartner WA, Hunter L and Verspoor K*. (2010) Test suite design for biomedical ontology concept recognition systems, in 7th Intl Conference on Language Resources and Evaluation.
63. Verspoor K*, Baumgartner W, Roeder C, Hunter L. (2009) Abstracting the Types away from a UIMA Type System, in Chiaros C, Eckhart de Castilho R, Stede M. (eds), Von der Form zur Bedeutung: Text automatisch verarbeiten / From Form to Meaning: Processing Texts Automatically. Tuebingen: Narr, 2009, p. 249-256.
64. Joslyn C, Paulson P, Verspoor K. (2008) Exploiting Term Relations for Semantic Hierarchy Construction. 2008 IEEE International Conference on Semantic Computing, pp.42-49.
65. Lin S-d and Verspoor K*. (2008). A Semantics-Enhanced Language Model for Unsupervised Word Sense Disambiguation. In Proceedings of the 9th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2008). Lecture Notes in Computer Science, vol. 4919, p. 287-298.
66. Joslyn C, Verspoor K, Gessler D. (2007). Knowledge Integration in Open Worlds: the Mathematics of Hierarchical Structure. In proceedings of First IEEE Int. Conf. on Semantic Computing.
67. Maguitman A, Rechtsteiner A, Verspoor K, Strauss CE, Rocha L. (2006) Large-Scale Testing Of Bibliome Informatics Using Pfam Protein Families. Pacific Symposium of Biocomputing 11:76-87
68. Dale R, Green SJ, Milosavljevic M, Paris C, Verspoor C, and Williams S (1998). Dynamic Document Delivery: Generating Natural Language Texts on Demand. Presented at DEXA'98.
69. Wan S and Verspoor C*. (1998), Automatic English-Chinese name transliteration for development of multilingual resources. In proceedings of COLING-ACL'98, Montreal, Canada.
70. Dale R, Green S, Milosavljevic M, Paris C, Verspoor C, and Williams S. (1998). Using Natural Language Generation Techniques to Produce Virtual Documents, In proceedings Australian Document Computing Symposium, ADCS'98.
71. Dale R, Green S, Milosavljevic M, Paris C, Verspoor C, Williams S. (1998). The Realities of Generating Natural Language from Databases. In proceedings 11th Australian Joint Conference on Artificial Intelligence.
72. Verspoor CM. (1998) The author explores the issue onto the agenda: Predictivity vs. Stipulativity in the Lexicon. In proceedings Pacific Asia Conference on Language, Information, and Computation, pp. 152-162.
73. Verspoor CM. (1996a) A Perspective on PPs. Edinburgh Working Papers in Cognitive Science, Vol. 12: Studies in Head-Driven Phrase Structure Grammar, Claire Grover and Enric Vallduvi, eds, pp. 229-271.
74. Verspoor CM. (1996b) Lexical Limits on the Influence of Context. In proceedings Eighteenth Annual Conference of the Cognitive Science Society. Lawrence Erlbaum Associates Publishers, pp. 116-120.

PEER-REVIEWED WORKSHOP/SIG MEETING PUBLICATIONS

75. Verspoor K, Kim JD, Dumontier M. (2015) Interoperability of Text Corpus Annotations with the Semantic Web. In proceedings, Biomedical Linked Annotation Hackathon (BLAH2015).
76. Choi JY, Verspoor K, Zobel J Analysis of Coreference Relations in the Biomedical Literature. (2014) Proceedings of the Australasian Language Technology Association Workshop, Melbourne, Australia, pages 134-138.
77. Groza T, Verspoor K Automated Generation of Test Suites for Error Analysis of Concept Recognition Systems. (2014) Proceedings of the Australasian Language Technology Association Workshop, Melbourne, Australia. pages 23-31.
78. Kocbek S, Verspoor K, Buntine W Exploring Temporal Patterns in Emergency Department Triage Notes with Topic Models. (2014) Proceedings of the Australasian Language Technology Association Workshop, Melbourne, Australia, pages 113-117.
79. Estival D, Cassidy S, Verspoor K, MacKinlay A, Burnham D. (2014) Integrating UIMA with Alveo, a human communication science virtual laboratory. In the Workshop on Open Infrastructures and Analysis Frameworks for HLT at the 25th International Conference on Computational Linguistics (COLING 2014), Dublin.
80. Pang C, Verspoor K, Chang S, Pearce J. (2014) Designing for Health Exploratory Seeking Behaviour. In Proceedings of the MedIR workshop at SIGIR'14.
81. Choi M, Verspoor K, Zobel J. (2014) Evaluation of Coreference Resolution for Biomedical Text. In Proceedings of the MedIR workshop at SIGIR'14.

KARIN M. VERSPOOR

82. Kahanda I, Funk C, Ullah F, Verspoor K, and Ben-Hur A. (2014) PHENOstruct: Prediction of human phenotype ontology using heterogeneous data sources. In Proceedings of the Automated Function Prediction SIG at ISMB'14.
83. Funk C, Kahanda I, Ben-Hur A and Verspoor K*. (2014) Evaluating a variety of text-mined features for automatic protein function prediction. In Proceedings of the BioOntologies SIG at ISMB'14.
84. Bada M, Baumgartner WA, Funk C, Hunter L, Verspoor K*. (2014) Semantic Precision and Recall for Concept Annotation of Text. In Proceedings of the BioOntologies SIG at ISMB'14. (*best paper award*)
85. Kahanda I, Funk C, Ullah F, Verspoor K, and Ben-Hur A. (2014) GOstruct 2.0: A framework for large-scale prediction of protein function. In Proceedings of the Automated Function Prediction SIG at ISMB'14.
86. Shmanina T, Zukerman I, Cavedon L, Jimeno Yepes A, Verspoor K*. (2013) Impact of Corpus Diversity and Complexity on NER Performance. In Proceedings of Australasian Language Technology Association Workshop, pages 91–95.
87. Matykiewicz P, Cohen KB, Holland KD, Glauser TA, Standridge SM, Verspoor KM, Pestian J. (2013) Earlier Identification of Epilepsy Surgery Candidates Using Natural Language Processing. In Proceedings of the BioNLP Workshop at the Association for Computational Linguistics 2013 meeting.
88. Jimeno Yepes A, Verspoor K*. (2013). Towards automatic large-scale curation of genomic variation: improving coverage based on supplementary material. In Proceedings of the BioLINK Special Interest Group Meeting at the Intelligent Systems for Molecular Biology 2013 meeting.
89. MacKinlay A, Martinez D, Jimeno Yepes A, Liu H, Wilbur WJ, Verspoor K*. (2013) Extracting Biomedical Events and Modifications Using Subgraph Matching with Noisy Training Data. In Proceedings of the BioNLP Shared Task Workshop at the Association for Computational Linguistics 2013 meeting.
90. Liu H, Verspoor K, Comeau DC, MacKinlay A, Wilbur WJ. (2013) Generalizing an Approximate Subgraph Matching-based System to Extract Events in Molecular Biology and Cancer Genetics. In Proceedings of the BioNLP Shared Task Workshop at the Association for Computational Linguistics 2013 meeting.
91. MacKinlay A and Verspoor K*. (2013) Information Extraction from Medication Prescriptions Within Drug Administration Data. The 4th International Workshop on Health Document Text Mining and Information Analysis with the Focus of Cross-Language Evaluation (LOUHI).
92. MacKinlay AD, Verspoor K*. (2012) Extracting Structured Information from Free-Text Medication Prescriptions, October 29, 2012. ACM Sixth International Workshop on Data and Text Mining in Biomedical Informatics (DTMBIO). Hawaii, USA.
93. Verspoor K*, Livingston K. (2012) Towards Adaptation of Linguistic Annotations to Scholarly Annotation Formalisms on the Semantic Web. In the Linguistic Annotation Workshop (LAW IV) at the Association for Computational Linguistics annual meeting.
94. Verspoor KM*, Cohn JD, Ravikumar KE, Wall ME. (2011) Integration of Structure Analysis and Text Mining for Improved Protein Functional Site Prediction. In Automated Function Prediction 2011 SIG at the Intelligent Systems for Molecular Biology meeting.
95. Sokolov A, Graim K, Funk C, Verspoor K, Ben-Hur A (2011) Combining Heterogeneous Data Sources for Protein Function Prediction. In Automated Function Prediction 2011 SIG at the Intelligent Systems for Molecular Biology meeting.
96. Livingston KE, Bada M, Hunter LE, and Verspoor KM* (2011) An Ontology of Annotation Content Structure and Provenance. In BioOntologies 2011 SIG at the Intelligent Systems for Molecular Biology meeting.
97. Liu H, Ravikumar KE, Verspoor K* (2011). From Graphs to Events: A Subgraph Matching Approach for Information Extraction from Biomedical Text. Proceedings of the 2011 Workshop on Biomedical Natural Language Processing, ACL-HLT 2011, pages 164–172.
98. Cohen KB, Christiansen T, Baumgartner Jr WA, Verspoor K, Hunter LE (2011). Fast and simple semantic class assignment for biomedical text. Proceedings of the 2011 Workshop on Biomedical Natural Language Processing, ACL-HLT 2011, pages 38–45.
99. Cohen KB, Johnson HL, Verspoor K, Roeder C. (2010). Tool performance and semantic type distribution for genotype-phenotype-drug relationship mining from text. In GPD-Rxn Workshop proceedings, Pacific Symposium of Biocomputing, January 2010, p. 3.
100. Ramakrishnan, C. Baumgartner WA, Blake, J.A, Burns, GAPC, Cohen KB, Drabkin H, Eppig J, Hovy E, Hsu C, Hunter LE, Ingulfsen T, Onda H, Pokkunuri S, Riloff E, Roeder, C and Verspoor K*. (2010). Building the Scientific Knowledge Mine (SciKnowMine): a community-driven framework for text mining tools in direct service to biocuration. In Proceedings of the NLP Frameworks Workshop at the Language Resources and Evaluation Conference (LREC), pp. 9-14.
101. Verspoor K*, Sanfilippo A, Elmore M, MacKerrow E. (2006). Deploying Natural Language Processing for Social Science Analysis. In proceedings Chicago Colloquium on Digital Humanities and Computer Science.

KARIN M. VERSPOOR

102. Joslyn C, Gessler, D.D.G, Schmidt, S.E, Verspoor KM. (2006) Distributed Representations of Bio-Ontologies for Semantic Web Services. In proceedings Joint BioLink and BioOntologies SIG at ISMB 2006.
103. Verspoor K*, Cohn J, Mniszewski S, Joslyn C. (2005). POSOLE: Automated Ontological Annotation for Function Prediction. In proceedings, Automated Function Prediction SIG at ISMB 2005.
104. Joslyn C, Cohn J, Verspoor K, Mniszewski S. (2005) Automating Ontological Function Annotation: Towards a Common Methodological Framework. In proceedings, BioOntologies SIG at ISMB 2005.
105. Verspoor C*, Joslyn C, and Papcun G (2003a). Interactions Between the Gene Ontology and a Domain Corpus for a Biological Natural Language Processing Application. In proceedings Sixth Annual Bio-Ontologies SIG Meeting at ISMB 2003.
106. Verspoor C*, Joslyn C, and Papcun G (2003b). The Gene Ontology as a Source of Lexical Semantic Knowledge for a Biological Natural Language Processing Application. In proceedings SIGIR'03 Workshop on Text Analysis and Search for Bioinformatics.
107. Verspoor C*, Dale R, Green S, Milosavljevic M, Paris C, and Williams S. (1998) Intelligent Agents for Information Presentation: Dynamic Description of Knowledge Base Objects. In proceedings International Workshop on Intelligent Agents on the Internet and Web, Mexico City, Mexico, pp. 75-86.
108. Verspoor CM*. (1997) Conventionality-Governed Logical Metonymy. In proceedings 2nd International Workshop on Computational Semantics, H. Bunt, L. Kievit, R. Muskens and M. Verlinden, eds, pp. 300-312.

OTHER PUBLICATIONS, ABSTRACTS, AND POSTERS

109. Verspoor K. (2015) Drawing on Millions of Biomedical Journal Publications to do Predictive Biology. Invited submission, 2015 International Conference on Big Data and Smart Computing (BigComp 2015).
110. Verspoor K. Practice-based evidence in medicine: Where information retrieval meets data mining. Invited submission, Medical Information Retrieval Workshop at SIGIR 2014, p. 4.
111. Funk C, Hunter L, Cohen K, Verspoor K*. (2014) Compositional rules for Gene Ontology synonyms generation. Poster presentation, Rocky Mountain Bioinformatics Conference.
112. Vassiliou T, Jimeno Yepes A, Wynne R, Gertz M, Knott K, Verspoor K*. (2014) Improving consistency of emergency department triage categorisation: Machine learning applied to clinical notes. Abstract presented at the Nursing Informatics Australia Conference at the Health Informatics Conference.
113. Martinez D, Bain C, Cavedon L, Verspoor K*. (2014) Text mining for lung cancer cases over a large collection of radiology reports. Abstract presented at the Health Informatics Society of Australia Big Data in Health and Medicine conference, Melbourne, April 2014.
114. Romano S, Bailey J, Cavedon L, Morrissey O, Slavin M, Verspoor K. (2014) Enhancing Diagnostics for Invasive Aspergillosis using Machine Learning. Abstract presented at the Health Informatics Society of Australia Big Data in Health and Medicine conference, Melbourne, April 2014.
115. Verspoor K*, Nicholson J. (2013) e-Learning with Kaggle in Class: Adapting the ALTA Shared Task 2013 to a Class Project. In Proceedings of Australasian Language Technology Association Workshop, pages 142–145.
116. Choi M, Jimeno Yepes A, Zobel J, Verspoor K*. (2013) NEROC: Named Entity Recognizer of Chemicals. BioCreative IV Workshop, Bethesda, MD.
117. Jimeno Yepes A, Neves M, Verspoor K*. (2013) Brat2BioC: conversion tool between brat and BioC. BioCreative IV Workshop, Bethesda, MD.
118. MacKinlay A, Verspoor K*. (2013) A Web Service Annotation Framework for CTD Using the UIMA Concept Mapper. BioCreative IV Workshop, Bethesda, MD.
119. Jimeno Yepes A, Verspoor K*. (2013) An assessment of the recovery of curated genetic variants through text mining. Abstract accepted for podium presentation in the Late Breaking Research track of ISMB/ECCB 2013.
120. Verspoor K*, Jimeno Yepes A; Ong C-S; Macintyre G (2013) Prioritising genetic mutations by mining the biomedical literature. Abstract presented at the Health Informatics Society of Australia Big Data in Health and Medicine conference, Melbourne 18-19 April 2013.
121. Cavedon L; Martinez D; Suominen H; Ananda-Rajah M; Pitson G; Verspoor K* (2013) Roles for language technology and text mining for next-generation healthcare. Abstract presented at the Health Informatics Society of Australia Big Data in Health and Medicine conference, Melbourne 18-19 April 2013.
122. Funk C, Baumgartner Jr W, Roeder C, Garcia B, Cohen KB, Hunter L, Verspoor K*. (2012) "Automated Concept Annotation: how are we doing?" Presented/Poster at Tenth Rocky Bioinformatics Conference.
123. Martinez DM, MacKinlay A, Molla-Aliod D, Cavedon L, Verspoor K*. (2012) Simple similarity-based question answering strategies for biomedical text, Sept 17-20, 2012. Conference and Labs of the Evaluation Forum (CLEF). Rome, Italy.

KARIN M. VERSPOOR

124. Funk C, Sokolov A, Ben-Hur A, [Verspoor K*](#) (2011) Improving Automated Protein Function Prediction by integrating Natural Language Processing and Machine Learning. Poster at Ninth Rocky Mountain Bioinformatics Conference.
125. Liu H, [Verspoor K*](#), Exploring an Approximate Subgraph Matching Approach for Biomedical Event Extraction. (2011) Poster at the Ninth Annual Rocky Mountain Bioinformatics Conference, Snowmass Village, CO, USA.
126. Sokolov A, Graim K, Funk C, [Verspoor K](#), Ben-Hur A (2011) Accurate Prediction of Protein Function Using GOstruct. Poster at the Critical Assessment Of Functional Annotation (CAFA) SIG meeting, ISMB 2011.
127. [Verspoor KM*](#), Cohn JD, Ravikumar KE, Wall ME (2011). Integration of Structure Analysis and Text Mining for Improved Protein Functional Site Prediction. In Automated Function Prediction 2011 SIG at the Intelligent Systems for Molecular Biology meeting.
128. [Verspoor K*](#) (2011) Annotating the Biomedical Literature with OBO and Linguistic Concepts. Use case submitted to the Using the OAC Model for Annotation Interoperability Workshop.
129. [Verspoor K*](#), Cohn J, Roeder C, Wall M. (2010) Validation of Protein Functional Site Predictions Using Automated Biomedical Literature Analysis. In Proceedings of the Eighth Rocky Bioinformatics Conference.
130. Bada M, Eckert M, Lanfranchi A, Baumgartner WA, Warner C, Howard A, Corvey W, Xue N, Cohen KB, [Verspoor K](#), Blake JA, Palmer M, Hunter L. (2010) Semantic Richness in the Colorado Richly Annotated Full-Text (CRAFT) Corpus. In Proceedings of the Eighth Rocky Bioinformatics Conference.
131. Roeder C, Christiansen T, Johnson HL, [Verspoor K](#), Burns G, Hunter L. (2010) Full-Text Biomedical Literature Processing: More than a Scaling Challenge. In Proceedings of the Eighth Rocky Bioinformatics Conference.
132. [Verspoor K*](#), Livingston, K.M, Roeder C, Christiansen T, Johnson HL, Cohen KB, Baumgartner WA, Hunter LE. (2010) The Colorado BioCreative III Gene Normalization task submission. Proceedings of the BioCreative III workshop, Bethesda, MD, September 13-15, 2010.
133. [Verspoor K*](#), and Mejia Muñoz C. (2010) Text Mining for Protein Function Prediction: Detection of Active Residues in Full-text publications. Poster presentation at ISMB 2010 and BioCreative III Workshop.
134. Livingston K, [Verspoor K](#), Hunter L. (2010) Improving Biomedical Text Mining by using Existing Background Knowledge, Poster presentation at ISMB 2010.
135. Cohen KB, Roeder C, Baumgartner WA, Hunter LE, [Verspoor K*](#). (2010) A test suite for ontology concept recognition systems: The Gene Ontology. Poster presentation at ISMB 2010.
136. Roeder, C. and [Verspoor K*](#). (2010) Scaling Text Mining to One Million Documents. Poster at ISMB 2010.
137. [Verspoor K*](#), Cohen KB, Johnson H, Roeder C, Mejia C, Baumgartner WA, Hunter L. (2009). Experiments with Biological Concept Recognition Tools. In Proceedings of the Seventh Rocky Mountain Bioinformatics Conference, p. 32.
138. Cohen KB, [Verspoor K](#), Roeder C, Baumgartner WA, Hunter L. (2009). Test Suite Design for Biomedical Ontology Concept Recognition Systems. In Proceedings of the Seventh Rocky Mountain Bioinformatics Conference, p. 33.
139. Cohen KB*, [Verspoor K*](#), Johnson H, Roeder C, Ogren P, Baumgartner W, White, E. Tipney H, and Hunter L. (2009) High-precision biological event extraction with a concept recognizer. In Proceedings of the Workshop on BioNLP: Shared Task, pages 50-58, Boulder, CO, June 2009.
140. [Verspoor K*](#), Roeder C, Johnson H, Cohen KB, Baumgartner WA, Hunter L. (2009). Information Extraction of Normalized Protein Interaction Pairs Utilizing Linguistic and Semantic Cues. In BioCreative II.5 Workshop Proceedings, p. 37.
141. Joslyn C, Gregory M, McGrath L, Paulson P, [Verspoor K](#). Semantic Hierarchies: Induction, Measurement, and Management. Poster at NSF Symposium on Semantic Knowledge Discovery, Organization and Use, NYU, November 2008.
142. [Verspoor K*](#), Baumgartner WA, Cohen KB, Johnson H, Hunter L. The Colorado OpenDMAP system: Building on Community Ontologies and a Community Platform for Biomedical Natural Language Processing. Poster at NSF Symposium on Semantic Knowledge Discovery, Organization and Use, NYU, November 2008.
143. Abi-Haidar A, Kaur J, Maguitman A, Radivojac P, Retchsteiner A, [Verspoor K](#), Wang Z, Rocha LM. (2007). Uncovering Protein-Protein Interactions in the Bibliome, in proceedings BioCreative 2006 Workshop.
144. [Verspoor K*](#), Joslyn C, Ambrosiano J (eds) (2005). Knowledge Integration for Bio-Threat Response. Los Alamos Unclassified Report LAUR-050907.
145. Narasimhan B, Di Tomaso V, and [Verspoor CM](#). (1996) "Unaccusative or Unergative? Verbs of Manner of Motion". Quaderni del laboratorio di linguistica, 10, Scuola Normale Superiore of Pisa.

KARIN M. VERSPOOR

PATENTS

- Australia Provisional Specification (2013) *Identification of Candidate Genetic Variants*. Verspoor CM, Macintyre G, Ong C-S, Jimeno Yepes A.
- (US Patent 7,933,856) (2011) *System and method for knowledge based matching of users in a network*. Verspoor CM, Sims BH, Ambrosiano JJ, Cleland TJ.

PROFESSIONAL SERVICE

- Associate Editor, *BMC Bioinformatics*, 2013-
- Scientific Programme Committee Chair, Health Informatics Society of Australia *Big Data* Conference, 2014, 2015
- Program Co-Chair IEEE International Conference on Machine Learning Applications (ICMLA) 2015
- President, Australasian Language Technology Association, 2013-2014
- Co-Chair, Australasian Language Technology Association Workshop, 2013
- Organizing Committee BioOntologies workshop at Intelligent Systems for Molecular Biology (ISMB), 2015
- Organizing Committee BioLINK workshop at Intelligent Systems for Molecular Biology (ISMB), 2014
- Chair, Organizing Committee BioLINK workshop at Intelligent Systems for Molecular Biology (ISMB), 2013
- NICTA representative on the Bio21 Cluster Scientific Advisory Council, 2012-2014
- Representative on the Alveo (formerly: Human Communication Science virtual lab) Steering Committee, 2013-
- Member, W3C Community Group on Open Annotation, 2012-
- NICTA representative, W3C Interest Group on Semantics in Healthcare and the Life Sciences, 2013-2014
- Grant Reviewer, US National Institutes of Health, 2010-2011
- Grant Reviewer, US Environmental Protection Agency, 2010-2011
- Grant Reviewer, US National Science Foundation, 2007-2012
- Member, Organizing Committee UCLA Institute for Pure and Applied Mathematics (IPAM) “Mathematics of Knowledge and Search Engines” program, Fall 2007
- Chair, UCLA IPAM Workshop on “Dynamic Searches and Knowledge Building”, October 2007
- Member, Organizing Committee UCLA IPAM “Social Data Mining and Knowledge Building” workshop, November 2007
- Member, OASIS-Open “Unstructured Information Management Architecture” standards committee, October 2006-2008.
- Chair, BioNLP’06. Linking Natural Language Processing and Biology, Workshop at NAACL 2006
- Los Alamos National Laboratory, Exploratory Research Grants Computer Science review committee, 2006
- Programme Committees and reviewing for numerous conferences, workshops, and journals
 - Journals (*inter alia*): PeerJ, PLoS One, Bioinformatics, BMC Bioinformatics, Briefings in Bioinformatics, Journal of Biomedical Informatics, Journal of Biomedical Semantics, Genome Biology, Journal of Biomedical Discovery, PLoS Computational Biology, Journal of the American Medical Informatics Association, IEEE/ACM Transactions on Computational Biology and Bioinformatics, SKY Journal of Linguistics, Journal of User-Modeling and User-Adapted Interaction, Applied Ontology, Information Processing and Management
 - Meetings (*inter alia*): Australian Language Technology Association Workshop, American Medical Informatics Association meeting, Australian AI Conference, Conference on Semantics in Healthcare and Life Sciences, Pacific Symposium on Biocomputing, Semantic Mining in Biomedicine, Language in Biomedicine, Data-Intensive Life Sciences, Health Informatics Society of Australia Big Data Conference, CICLING, BioCreative, Intelligent Systems for Molecular Biology (Text Mining session and BioLink SIG workshops), European Conference on Computational Biology, European Association for Computational Linguistics, Language in Biomedicine, Recent Advances in Natural Language Processing, Association for Computational Linguistics BioNLP Workshop, Knowledge Discovery in the Life Science Literature, IEEE Systems Man and Cybernetics, annual meeting of the Cognitive Science Society

GRANTS AND CONTRACTS AWARDED

- Australian Research Council Discovery Project (Chief Investigator), \$355k, 2015-2018
Natural language processing for automated validation of protein databases
- Victoria Department of Health (Principal Investigator), \$205k, 2014-2015
Biomedical Informatics for effective health decision making
- Defense Science and Technology Organisation, (Principal Investigator), \$75k, 2014
The Application of Natural Language Processing Techniques to Medical Information Systems for Syndromic Surveillance

KARIN M. VERSPOOR

- Andrew W Mellon Foundation, Award No. 2011-02048-05 (Principal Investigator), \$45k, 2011-2012
Adapting the OAC data model for automated annotation of the biomedical literature
- National Science Foundation, DBI-0965616 (Co-investigator), \$288k, 2010-2014
GOSTRUCT: modeling the structure of the Gene Ontology for accurate protein function prediction
- National Institutes of Health, National Library of Medicine, R01LM010120-02S1 (Principal Investigator), \$80k, 2010-2011. Automated Literature Mining for Validation of High-Throughput Function Prediction (Supplement)
- National Institutes of Health, National Library of Medicine, R01LM010120 (Principal Investigator), \$1.4M, 2009-2011. Automated Literature Mining for Validation of High-Throughput Function Prediction
- Los Alamos National Laboratory Exploratory Research Grant (Principal Investigator), \$775k, 2006-2008
Unsupervised methods for Induction of Patterns for Information Extraction
- Defense Threat Reduction Agency, Technology Innovation Consortium project, 2006-2008
- Department of Homeland Security, ADVISE program, 2005-2007
- Los Alamos National Laboratory-New Mexico State University collaboration program, 2005-2006
- Department of Homeland Security, Motivation and Intent program, 2006-2007
- Cooperative Research and Development Agreement, Proctor & Gamble Corporation, 2003-2004
- Macquarie University Research Grant (Principal Investigator), \$7k, 1997
- UK Engineering and Physical Sciences Research Council Research Studentship for PhD study, 1994-1997
- Doctor Catherine van Tussenbroek grant to fund PhD research visits to the University of Utrecht and Brandeis University, 1995-1996

DIVERSITY AND OUTREACH ACTIVITIES

- Panelist, RMIT University Society for Women in Information TeCHnology (SWITCH) event on Women in IT, September 2014
- Speaker, 2014 Victorian ICT for Women “Go Girl, Go for IT” event, “The Wandering Computer Scientist”
- Volunteer, 2014 Victorian ICT for Women “Go Girl, Go for IT” event (school liaison)
- Microsoft International Women’s Hackathon, Women in ICT at Uni Melbourne; 2014: Judge, 2013: speaker
- Robogals Science Challenge 2013, judge, 11/2013
- Panelist, Women in Technology panel at the "Wired for Wonder" conference in Sydney, 08/2013
- She++ documentary screening and panel discussion: organizer and panel moderator, 05/2013
- NICTA Diversity Council member, 2013-2014
- Primary school children instruction, “CS Unplugged” curriculum, Ebert Elementary School, Denver, Colorado 2010-2011
- Los Alamos National Laboratory, Women’s Diversity Working Group, co-chair 2006-2008
- Los Alamos National Laboratory, Diversity/Affirmative Action Board, Secretary 2005-2007