

Dr SARVNAZ KARIMI

Senior Research Scientist
Language and Social Computing Team
Analytics and Decision Sciences
CSIRO Data61
Marsfield NSW 2122, Australia
Phone: +61 2 9372 4353
Mobile: +61 413 179 097
Email: sarvnaz.karimi@csiro.au

Bio

Sarvnaz is a senior researcher at CSIRO's Data61. She has over 12 years of experience working in health informatics, applying her expertise at the intersection of Information Retrieval, Natural Language Processing (NLP), and Machine Learning.

Sarvnaz is very active in the research community at large, in Australia and internationally. Sarvnaz is currently deputy director of publicity at the Association for Computational Linguistics (ACL), which is the main international professional body for NLP. She is the president elect of Australasian Language Technology Association (ALTA). She has served as chair, area-chair, and programme committee member of a number of major national and international conferences, including Australasian Document Computing Symposium (ALTA), Australasian Document Computing Symposium (ADCS), the Association for Computing Machinery's Special Interest Group on Information Retrieval (SIGIR), The Conference on Information and Knowledge Management (CIKM), the International Conference on Computational Linguistics (COLING), and the Association for Computational Linguistics (ACL). She serves as reviewer for a number of journals including ACM Computing Surveys, Artificial Intelligence for Medicine, Artificial Intelligence in Medicine, BMC Medical Informatics and Decision Making, Language Resources and Evaluation, and Information Processing & Management. Through CSIRO and the Data61 scholarship programme, Sarvnaz currently co-supervises five PhD students from different Australian universities.

Sarvnaz's formal background is in Computer Science and Software Engineering. She completed a PhD in Natural Language Processing in 2008. She has a masters in Artificial Intelligence, and a bachelor's degree in Software Engineering.

Homepage: people.csiro.au/K/S/Sarvnaz-Karimi

LinkedIn: www.linkedin.com/in/sarvnaz-karimi-64ab491

Google Scholar (Publications): goo.gl/KoTmwz

Publons (Journal reviews): publons.com/a/1398968/

Research Interests

Digital Health, Biomedical Text Mining, Natural Language Processing, Search and Information Retrieval, Social Media Analytics.

Education

February 2005 – March 2008 PhD in Computer Science

School of Computer Science and Information Technology

RMIT University, Melbourne, Australia

Research topic: Machine transliteration of proper names between English and Persian.

Supervisors: Prof. Falk Scholer, Prof. Andrew Turpin.

The thesis explored the problem of English-to-Persian and Persian-to-English transliteration using statistical machine learning methods based on the graphemes of words. It proposed three novel approaches that use consonant-vowel sequences, and showed significant improvements over baseline systems. That work also contributed a novel text alignment algorithm and examined new techniques to combine systems using classifiers to improve the effectiveness of the systems. The contributions of this thesis are published in the Association for Computational Linguistics (ACL) Conferences (CORE Ranking A*) as well as in an article in the highly cited ACM Computing Surveys (Impact Factor: 6.131).

Professional Experience

Mar 2012 – present *Senior Research Scientist*

CSIRO Data61, Marsfield, NSW

- Research member of Language and Social Computing Team, Analytics and Decision Sciences Programme within Data61.
- Dec 2019—present: Chair of NSW/SA CSIRO Data61 PhD scholarship committee. It involves assessing all the scholarship applicants, chairing the assessment sessions and finalising a recommendation of the ranking of the applicants to the CSIRO Data61's chief scientist.
- Nov 2017—Nov 2019: Member of CSIRO Data61 PhD scholarship committee for VIC/TAS/WA states. This role involves reviewing PhD scholarship applications and ranking them in a committee meeting three to four times a year.

Projects:

- Syndromic Surveillance for Infectious Diseases (2017–present)
The goal of this project is to use social media, such as Twitter, to assist in identifying infectious disease in specific geographical locations. This project is in collaboration with the UNSW Public Health research group. We were successful to attract an Office of the Chief Executive (OCE) grant to hire a postdoctoral fellow for this project as well as funding from CSIRO Health & Biosecurity's Risk Evaluation and Preparedness Programme. As part of this project we have published 4 journal papers and 4 conference papers. Our team was ranked first in an international shared task on vaccination behaviour detection from social media. A tutorial on this area (Syndromic Surveillance) is also accepted at one of the leading computational linguistics conferences in December 2020.

- Medical Decision Support (2018–present)

Leading a Precision Health Future Science Platform project on improving decision making for precision medicine. We collaborate with researchers in a Belgium university to apply our literature search platform for proteomics. Our search system ranked among top 10 teams internationally in the TREC evaluation tasks on precision medicine in 2019.

We also tailored our search system for COVID search to improve accessing information in published literature on coronavirus as outlined in an international shared evaluation task. Our team ranked 4th internationally among 143 submissions.
- Low Carb Meals Commercialisation (2019–present)

Responsible for evaluating the effectiveness of the CSIRO’s low carb diet book series and the endorsed meals through social media feedback.
- Deep learning for Clinical Support (2018–2019)

In collaboration with researchers in the CSIRO’s eHealth programme, developed algorithms that allow processing clinical documents, such as radiology reports, across multiple hospitals. We collaborated with clinicians from Gold Coast hospitals to evaluate such transferrable systems.
- Search for Clinical Decision Making (2016–2018)

Led this project with two aims: (1) developing a public platform for researchers to experiment on information retrieval algorithms that advances search for biomedical articles; and, (2) developing algorithms that match cancer patients with clinical trials. This project is later transformed to a Precision Health Future Science Platform Project.
- Artificial Intelligence Roadmap (2018–2019)

Part of the team that wrote Artificial Intelligence Roadmap: Solving problems, growing the economy and improving our quality of life. This report is requested and delivered to the Australian Government.
- Information extraction for digital leasing (2017)

Providing consultancy and a proof-of-concept product, in the domain of information extraction, for an SME in Sydney.
- Australia At Your Service (2016)

An expert search system was developed in this project called Network (network.csiro.au), which is currently in use internally within CSIRO. This project also led to some involvement in another expertise search project called ExpertConnect (<https://expertconnect.data61.csiro.au/>).
- Adverse Drug Reaction Discovery from Patient Reports (2013–2015)

Led a two year project to (1) mine patient forums for reports of adverse drug events; and, (2) find duplicate reports of adverse events in the Therapeutic Goods Administration (TGA) spontaneous reporting system. This project led to highly cited publications in journals such as ACM Computing Surveys and Journal of Biomedical Informatics, as well as two software prototypes. We also released a dataset called CADEC through a highly cited journal paper (over 100 citations) that is internationally used for developing similar methods in the biomedical text mining community.

- Enhanced coding of cause of death recorded on death certificates (2015–2016)
Applied text classification methods on free-text reports of death certificates to automatically assign ICD10 codes (International Coding of Diseases). The outcome was a system that is currently under use in the NSW Health influenza monitoring system.
- Emergency Situation Awareness (2012–2013)
As a research member of the team provided NLP expertise and developed methods to identify social media posts that indicate different categories of disasters, as well as developing geolocation prediction methods for Twitter messages. During this time also co-supervised two interns and one honours thesis from the University of Sydney.

May 2008 – Mar 2012 *Researcher*

NICTA Victoria Research Laboratory, The University of Melbourne, VIC
Research member of Biological Text and Language Applications (BioTALA) project, Making Sense of Data (MSD) theme.

In BioTALA, developed techniques for processing biomedical text. Some of these activities were focused on specific needs of external collaborators (e.g., The Global Evidence Mapping Initiative and Alfred Hospital).

My main research centered around developing algorithms for improving search in biomedical text, including search for relevant published research for creation of clinical systematic reviews, and genomics information retrieval. Also, worked on extracting drug side-effects from public forums, and search in medical records.

Semester 2, 2011 *Guest Lecturer*

School of Computer Science and Information Technology
RMIT University, Melbourne, Australia

- Information retrieval

February 2005 – October 2007 *Sessional Tutor*

School of Computer Science and Information Technology
RMIT University, Melbourne, Australia

- Database Concepts (5 semesters)
- Web Database Applications (2 semesters)
- Scripting Language Programming (Perl and Python)
- File Structures and Database Systems (2 semesters), Head Tutor
- Web Programming
- Web Servers and Web Technology
- Computer and Internet Forensics
- Secure Electronic Commerce
- Unix Essentials and C shots (RMIT Orientation Programme, four semesters)

September 2005 – November 2005 *Online Instructor*
School of Computer Science and Information Technology
RMIT University, Melbourne, Australia
Introduction to Information Technology

August 1999 – January 2005 *Research Engineer*
Iran Telecommunication Research Centre (ITRC), Tehran, Iran

- Member of text mining research group (October 2002 – January 2005)
Responsibilities:
Improving the performance of TelQAS, an ontology-based telecommunication question answering system developed in-house, using query expansion (pseudo-relevance feedback), document clustering, and multi-document summarisation.
- Member of network management research group (August 1999 – September 2002)
Responsibilities:
Improving basic call service model in Intelligent Networks (IN)
Real-time services in Internet, Voice Over IP (VOIP)
TMN (Telecommunication Management Network) in mobile networks

Honours and Awards

Outstanding reviewer award from the Empirical Methods in Natural Language Processing (EMNLP) Conference in 2018 and 2019 (CORE ranking A conference in Natural Language Processing).

Part of the LaSC *LIS-Accurait* project team that won two merit awards at the 2019 NSW iAwards ceremony for *Infrastructure and Platforms Innovation of the Year* and *Business Service Markets*.

OCE Postdoctoral Fellowship grant on “Improving the detection, diagnosis and prognosis of infectious diseases by integrating current syndromic surveillance tools with social media insights”, with the value of \$105K per year for three years (2018–2021), together with Ross Spark, Cecile Paris, Paul De Barro, and Raina McIntyre (UNSW). This allows us to hire a postdoc under our supervision.

CSIRO non-cash award in recognition of praiseworthy achievement in supporting “The Power of Peoples Data” stand at D61+Live 2019.

CSIRO Program Recognition Rewards in Excellence, 2017. As part of COLING team that won an international competition to identify the location of twitter posts.

CSIRO Program Recognition Rewards in Performance, 2017. As part of NETWORK team which negotiated CSIROs social, institutional and technical complexity to develop and launch a product to improve CSIROs capability discoverability in a very short space of time (<http://network.csiro.au>).

Commendation on the PhD thesis to be recognised as a work that is *of the highest merit and at the forefront internationally in its (examiners: A/Prof James Thom, Prof. Kevin Knight, A/Prof Ari Pirkola). field*, by the international thesis examiners, 2008.

Best full paper award, The 14th Australasian Document Computing Symposium (ADCS'09), Sydney, Australia, December 2009.

Member of the team which became semi-finalist at the Elsevier Grand Challenge (from 70 entries): Timothy Baldwin, Lawrence Cavedon, Sarvnaz Karimi, David Martinez, David Newman, Falk Scholer and Justin Zobel, "Effective Search, Classification, and Visualisation of Information from Large Collections of Biomedical Literature".

Best presentation award, RMIT research student conference (CRSC), 2006.

Student travel grant (£500) to attend String Processing and Information Retrieval (SPIRE), 2006, Glasgow, UK.

Ranked 15th, the Iran trans-country university entrance exam (for masters degree), 2001.

Top student, gold coin recipient (2nd place among the university graduates of BEng, Computer Engineering), 2000.

University student research prize, 1999.

Member of the university ACM programming team which ranked 7th in Tehran site, West Asia region of the ACM international collegiate programming contest, 1999.

Other

Certificate of Appreciation, The 2014 ARCS Scientific Conference (for invited speech entitled *Text Mining for Pharmacovigilance*).

Certificate of Completion, Project Management Fundamentals, CSIRO, 2013.

Publications

A full list of publications and their citations can be found in my Google Scholar page: [goo.gl/K0Tmwz](https://scholar.google.com/citations?user=K0Tmwz). Some of my journal articles are listed below: (IF: Impact Factor)

- W. Tay, J. Zhang, S. Karimi, Beyond mean rating: Probabilistic aggregation of star ratings based on helpfulness, *Journal of the Association for Information Science and Technology*, pages 784-799, 71(7), 2020. (CORE ranking A*, IF: 2.835)
- H. Hassanzadeh, S. Karimi, A. Nguyen, Matching Patients to Clinical Trials Using Semantically Enriched Document Representation, *Journal of Biomedical Informatics*, 2020. (IF: 2.95)
- A. Joshi, R. Sparks, J. McHugh, S. Karimi, C. Paris, CR. MacIntyre, Harnessing tweets for early detection of an acute disease event, *Epidemiology*, 31 (1), 2020. (IF: 2.26)
- A. Joshi, S. Karimi, R. Sparks, C. Paris, CR. MacIntyre, Survey of text-based epidemic intelligence: A computational linguistic perspective, *ACM Computing Surveys*, 52(6), Article 119. (CORE ranking A*, IF: 6.131)
- B. Koopman, S. Karimi, A. Nguyen, R. McGuire, D. Muscatello, M. Kemp, D. Truran, M. Zhang, S. Thackway, Automatic classification of diseases from free-

text death certificates for real-time surveillance. *BMC Medical Informatics and Decision Making*, 15(53), 2015. (IF: 2.067)

- S. Karimi, J. Yin, J. Baum, Evaluation Methods for Statistically Dependent Text, *Computational Linguistics*, pages 539-548, 41(3), 2015. (CORE ranking A*, IF: 1.319)
- S. Karimi, A. Metke-Jimenez, M. Kemp, C. Wang, CADEC: A Corpus of Adverse Drug Event Annotations, *Journal of Biomedical Informatics*. pages 7381, Vol. 55, June 2015. (IF: 2.95)
- S. Karimi, C. Wang, A. Metke-Jimenez, R. Gaire, C. Paris, Text and Data Mining Techniques in Adverse Drug Reaction Detection, *ACM Computing Surveys*, Vol. 47, Number 4, Article 56, 2015 (CORE ranking A*, IF: 6.131)
- S. Karimi, J. Zobel, F. Scholer, Quantifying the Impact of Concept Recognition on Biomedical Information Retrieval, *Information Processing and Management*, pages 94-106, 48(1), 2012. (CORE ranking A, 3.892)
- S. Karimi, F. Scholer, A. Turpin, Machine Transliteration Survey, *ACM Computing Surveys*, 57 pages, 43(3), April 2011. (CORE ranking A*, IF: 6.131)
- S. Karimi, S. Pohl, F. Scholer, L. Cavedon, J. Zobel, Boolean versus Ranked Querying for Biomedical Systematic Reviews, *BMC Medical Informatics and Decision Making*, 10(58), 2010. (IF: 2.067)

Science Citizenship

Note: Computer science conferences, in general, provide peer-reviews for the submitted papers. Each *area chair* is in charge of a subset of submitted manuscripts for which they assign reviewers, moderate a discussion phase, and provide tentative decisions which are handed to technical program chairs.

Chair/Co-chair/Area chair:

- Area chair of EMNLP 2020 (The 2020 Conference on Empirical Methods in Natural Language Processing)—Machine Learning for NLP track
- Area chair of ACL 2020 (The 58th annual meeting of the Association for Computational Linguistics)—Resources and Evaluation track
- Area chair of ACL 2019 (The 57th annual meeting of the Association for Computational Linguistics)—NLP Applications track
- Area chair of COLING 2018 (The 27th International Conference on Computational Linguistics)—Information extraction tasks
- Publicity chair of ACL 2018 (The 56th Annual Meeting of the Association for Computational Linguistics)
- Area chair of ACL 2017 (The 55th Annual Meeting of the Association for Computational Linguistics)—Information Extraction and Retrieval, Question Answering, Text Mining, Document Analysis and NLP Applications track
- Co-chair of ADCS 2015 and 2016 (Australasian Document Computing Symposium).

- Co-chair of ALTA 2013 (Australasian Language Technology Association Workshop)

Steering/executive committee membership:

- Deputy Director of Publicity, Association for Computational Linguistics (2019-2021)
- Australasian Language Technology Association (2015-present)
- Australasian Document Computing Symposium (2016-present)

Journal Reviewer:

A complete list of reviews is available on Publons.

- Artificial Intelligence in Medicine (AIIM)
- BMC Medical Informatics and Decision Making
- Journal of Biomedical Informatics (JBI)
- Information Processing and Management (IPM)
- Journal of Information Sciences (INS)
- ACM Computing Surveys
- Journal of Language Resources and Evaluation (LRE)
- ACM Transactions on Asian Language Information Processing (TALIP)
- IEEE Transactions on Knowledge and Data Engineering (TKDE)
- Information Retrieval, Springer

Conference Program Committee or Reviewer:

- The 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)
- ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR 2021)
- The 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 9th International Joint Conference on Natural Language Processing (AAACL-IJCNLP 2020)
- NLP COVID-19 Workshop an emergency workshop at ACL 2020
- The Conference on Artificial Intelligence (AAAI 2020)
- The Annual Symposium of American Medical Informatics Association (AMIA 2019)
- The Annual Symposium of American Medical Informatics Association (AMIA 2018)
- The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, Demonstrations (SIGIR 2018)

- The 15th Annual Workshop of the Australasian Language Technology Association (ALTA 2017)
- The 22nd Australasian Document Computing Symposium (ADCS 2017)
- The 14th Annual Workshop of the Australasian Language Technology Association (ALTA 2016)
- The Sixth ACL Named Entities Workshop (NEWS 2016)
- The 24th ACM International Conference on Information and Knowledge Management (CIKM 2015)
- The 13th Annual Workshop of the Australasian Language Technology Association (ALTA 2015)
- The 10th International Persuasive Technology Conference (PT 2015)
- The 5th Named Entities Workshop in conjunction with ACL (NEWS 2015)*
- The 9th Conference on Information Technology in Asia (CITA 2015)
- The 37th International ACM SIGIR conference on Research & development in information retrieval (2014) Short paper, Demo, and Tutorial tracks
- The 19th Australasian Document Computing Symposium (ADCS 2014)
- The 12th Annual Workshop of the Australasian Language Technology Association (ALTA 2014)
- The 9th International Persuasive Technology Conference (PT 2014)
- The 36th International ACM SIGIR conference on Research & development in information retrieval (2013) Short paper
- The 18th ACM International Conference on Intelligent User Interfaces (IUI 2013)
- The 2nd Workshop on Advances in Text Input Methods in conjunction with COLING (WTIM 2012)
- The 4th Information Interaction in Context Symposium (Iiix), 2012.
- The Annual International ACM SIGIR Conference, posters, 2012.
- The 50th Annual Meeting of the association for computational linguistics (ACL), demonstrations, 2012.
- The Named Entities Workshop (NEWS) 2012, in conjunction with ACL, 2012.
- The 5th International Joint Conference on Natural Language Processing (IJCNLP), 2011.
- The Named Entities Workshop (NEWS) 2011, in conjunction with IJCNLP, 2011.
- The 6th Asia Information Retrieval Society Conference (AIRS), 2010.
- The 6th International Conference on IT in Asia (CITA), 2009.
- The Australasian Language Technology Workshop (ALTW), 2008.
- Reviewer for The Australasian Database Conference, ADC 2008.
- Reviewer for Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL-HLT 2007.

- Reviewer for The Annual International ACM SIGIR Conference, poster, SIGIR 2007.
- Reviewer for String Processing and Information Retrieval, SPIRE 2006.
- Reviewer for Australasian Language Technology Workshop, ALTW 2006.
- Reviewer for The International Conference on Information Integration and Web Based Applications Services, IIWAS 2005.

Organising committee member of RMIT Computer Science and IT Research Student Conference (CRSC), 2006.

Invited Speaker, Visiting Scientist, Media

- Media report by Medical Republic, January 2020
An interview with a journalist led to an article *Why health is the 'worst case' scenario for text mining*.
- Tutorial Speaker at Australasian Language Technology Association 2019 Tutorial, December 2019
Title: *NLP for Healthcare in the Absence of a Healthcare Dataset*, co-presented with Aditya Joshi.
- Speaker at CSIRO Connect 2019, August 2019
Title: *AI for Health*.
- Speaker for Big Day of Data61 Scholars, February 2019
Title: *Your social media and our research*.
- Panelist at Sydney Natural Language Processing Meetup, April 2018
Panel topic: *The next 30 years of information extraction*.
- Invited speaker at CMCRC (Capital Markets CRC), July 2017
Title: *Information extraction and retrieval for decision making in health and precision medicine*.
- Invited talk at Australian Institute of Health and Welfare, 2017
Title: *Textual Analysis: Techniques, New Data , Opportunities and Challenges*, co-presented with Cecile Paris.
- Invited talk at PwC (PriceWaterhouseCoopers) Australia, 2015
Presented on behalf of Scientists & Mathematicians in School to share my experience as a volunteer scientist.
- Invited speaker at ARCS Australia Scientific Congress (<https://www.arcs.com.au/>), ARCS Pharmacovigilance and Social Media, June 2014
Title: *Text Mining for Pharmacovigilance*, presented at Pharmacovigilance and Social Media Session.
- Invited talk at Amsterdam University, Information Retrieval Group, 2011
Title: *Systematic Reviews for Evidence Based Medicine*.
- Visiting scholar at Glasgow University, Information Retrieval Group, November, 2011.

Volunteering:

- Scientist in School (2012–present). In partnership with teachers in schools, teaching some core topics in computer science (e.g., algorithms) and giving presentations to encourage students to study computer science/engineering.
- Diversity and women representative officer of NICTA (National ICT Australia), Victoria research laboratory (2008–2012).

Student Supervision

2019—present, Vincent Nguyen, PhD Student on Question Answering, Australian National University (co-supervisor: Zhenchang Xing)

2018—present, Wenyi Tay, PhD Student on Opinion Summarisation, RMIT University (co-supervisors: Jenny Zhang and Stephen Wan)

2018—present, Sonit Singh, PhD student on Radiology Report Generation, Macquarie University (co-supervisors: Len Hamey, Kevin Ho-Shon)

2017—present, Xiang Dai, PhD Student on Information extraction, Sydney University (co-supervisors: Cecile Paris, Ben Hachey)

2016—2017, Xiang Dai, Industrial trainee, University of Sydney

2016—2018, Vincent Nguyen, CSIRO summer student, University of Sydney

2014—2015, Han Zhao, CSIRO Summer student, University Technology of Sydney

2013—2013, Allan Jun Yin, Industrial trainee, Australian National University

2012—2013, John Lingad, CSIRO summer student and honours degree, University of Sydney