David Vadas

Mobile: +44 7435 159 812
Email: dvadas@gmail.com
GitHub: https://github.com/dvadas

Employment History

Google, as an Adecco contractor **Jan 2015 – Present**

I'm currently working in the Text-To-Speech (TTS) team, developing a major new voice for Google's production systems. This involves preparing data for the recording process, and ensuring the quality of lines synthesised by the voice. My work on this new voice has resulted in measurably improved performance on the lines that are most important to our users.

A key achievement during my time here has been to automate many tasks that were previously done manually. The team has multiple data sources that need to be kept synchronised, and removing the repetitive human effort to do so is a great help for many people on the TTS team.

I have gained a lot of experience working with Google technology in this position. I work extensively with BigTables, both programmatically and from the command line. I have written many unit tests with the gUnit framework; in one instance I reduced test time by so much that a "long" timeout could be removed.

Susquehanna **Sep 2011 – Oct 2014**

I worked on a specialist high-frequency trading desk that ran strategies without full-time trader supervision. During my time there I developed and optimised a brand new trading strategy that was profitable in a backtest environment. My strategy also provided many insights that were applied in other trading models.

I also worked on the trading and backtesting infrastructure, writing key improvements to the core system. This complex C++ code would be used throughout the group environment and by all our strategies. I developed connectivity code to receive tick data from the exchange, and wrote a data processing system to transform the full book tick data into key predictors that were used by the trading strategies.

University of Sydney and Capital Markets CRC **Nov 2010 – Aug 2011**

I worked on an entity linking system that automatically generated content pages from newspaper articles. My key contribution in this position was to design and implement a new database schema to store all the information extracted from our corpus. The system has now been deployed by Fairfax Media, one of the largest media companies in Australia.

I lectured a Natural Language Processing course and supervised students.

Optiver **Apr 2008 – Nov 2010**

I managed a data capture and analysis system distributed across multiple geographic locations that processed huge volumes of data. The system provided submicrosecond timing information on our trading strategies, letting the company know where it was most valuable to spend programmer effort.

I also worked on a specialised high-frequency trading strategy, increasing the profit that it made.

University of Sydney **2003 – 2006**

I tutored for many programming courses, from high school students through to honours and masters students.

Education

The University of Sydney
PhD in Natural Language Processing **Mar 2005 – Apr 2008**
Thesis: *Statistical Parsing of Noun Phrase Structure*
Bachelor of Information Technology (Honours) **Mar 2001 – Nov 2004**
First Class Honours in Computer Science – Grade: 88/100 (WAM: 80%)
Majors in Software Development and Principles of Computer Science.

Technical Skills

- Languages: C++ (expert), Python (expert), C (proficient).
- Extensive practical experience working with large-scale data, including applying machine learning techniques, analysis with numpy and scipy, and writing code to run on distributed systems.
- Demonstrated ability with Google systems such as BigTable, gUnit, and Google's C++ library.
- Considerable practice at using Linux utilities, e.g. awk, to speed prototyping work.
- Knowledge of how to design efficient database schemas and write complex SQL queries.
- Domain-specific experience in high frequency trading and natural language processing.
- Excellent communication skills from lecturing, writing technical papers, and working with teammates.

Publications

David Vadas and James R. Curran
Parsing Noun Phrases in the Penn Treebank. In Computational Linguistics, 37(4), pages 753–809. December 2011.

David Vadas and James R. Curran
Parsing Noun Phrase Structure with CCG. In Proceedings of the 46th Annual Meeting of the Association of Computational Linguistics: Human Language Technologies (ACL-08: HLT). Columbus, OH, USA, June 15–20 2008.

David Vadas and James R. Curran
Parsing Internal Noun Phrase Structure with Collins' Models. In Proceedings of the Australasian Language Technology Workshop (ALTW-07), pages 109–116. Melbourne, Australia, December 10–11 2007.

David Vadas and James R. Curran
Large-Scale Supervised Models for Noun Phrase Bracketing. In Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics (PACLING-2007), pages 104–112. Melbourne, Australia, September 19–21 2007.

David Vadas and James R. Curran
Adding Noun Phrase Structure to the Penn Treebank. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL-07), pages 240–247. Prague, Czech Republic, June 23–30 2007.

James R. Curran, Stephen Clark, and **David Vadas**
Multi-Tagging for Lexicalized-Grammar Parsing. In Proceedings of the Joint Conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics (COLING/ACL-06), pages 697–704. Sydney, Australia, July 17–21 2006.

David Vadas and James R. Curran
Tagging Unknown Words with Raw Text Features. In Proceedings of the Australasian Language Technology Workshop (ALTW-05), pages 32–39. Sydney, Australia, December 10–11 2005.

David Vadas and James R. Curran
Programming With Unrestricted Natural Language. In Proceedings of the Australasian Language Technology Workshop (ALTW-05), pages 191–199. Sydney, Australia, December 10–11 2005.