SYDNEY NLP MINI CONFERENCE #1

JULY 14TH, 2018

13:00 — OPENING & WELCOME

The [Australasian Language Technology Association](#) and [Sydney NLP Meetup](#) welcome you to the Sydney NLP Mini Conference, hosted by [Tyro FinTechHub](#).

We wish to acknowledge the custodians of this land, the Gadigal people of the Eora nation and their Elders past and present. We acknowledge and respect their continuing culture and the contribution they make to the life of this city and this region.

13:15 — MATT HONNIBAL / EXPLOSION AI

spaCy's next parsing model: A non-monotonic transition-based approach to joint segmentation and parsing

ABSTRACT

spaCy is a popular open-source Natural Language Processing library designed for practical usage. In this talk, I'll outline the new parsing model we've been developing to improve spaCy's support for more languages and text types. Like other transition-based parsers, the model predicts a sequence of actions that push tokens to and from a stack and build arcs between them. However, we expect the arc-eager system with actions that can also repair previous parse errors, introduce sentence boundaries, and split or merge the pre-segmented tokens. The joint approach improves parse accuracy on many types of text, especially for non-whitespace writing systems. We have also found significant practical advantage to short pipelines. Short pipelines are easier to reason about, and increase runtime flexibility by reducing the risk of train/test skew.

BIOGRAPHY

Matthew Honnibal began his research career as a linguist, completing his PhD in 2009 on lexicalised parsing with Combinatory Categorial Grammar, before working on incremental speech parsing with Mark Johnson. These days he is best known as a software engineer, for his work on the spaCy NLP library. He grew up in Sydney, lives in

Berlin, and still misses CCG.

13:45 — RAHUL GOEL / AMAZON ALEXA

Improving conversational interactions @Alexa

ABSTRACT

For a service like Alexa, context comes in various forms: past user utterances, user information, device state, screen information etc. In this talk we discuss how we can incorporate context, specifically screen context and past conversational context, to make Alexa more conversational. We will also touch upon how we can judge the quality of non-goal oriented conversations using engagement and coherence as proxy metrics.

BIOGRAPHY

Rahul Goel is a Machine Learning scientist at Amazon Alexa where he works on developing language technologies for products like Amazon Echo and Amazon Echo Show. His research interests include Spoken Language understanding, dialog systems, deep learning, multi-modal modeling and social computing. Before joining Amazon he was a grad student at Georgia Tech working with Dr. Jacob Eisenstein on computational social science.

14:15 — BREAK

14:30 — INES MONTANI / EXPLOSION AI

Prodigy: an annotation tool designed for rapid iteration and developer productivity

ABSTRACT

Most developers working with machine learning recognise that data quality and quantity is a more important factor for the success of their project than the specifics of their statistical model. Despite this, it's common for inexperienced teams to make almost no investment into their data. Even amongst more experienced teams, developers often under-estimate the extent to which annotation is a knowledge-based process that requires several iterations to perfect. As a solution, we suggest machine learning developers perform initial annotations themselves, to help them refine the schema. To

enable this workflow, we've developed Prodigy, an annotation tool with several features designed to improve productivity. In this talk I'll discuss what we've learned about annotation, and show you how we've implemented these insights into Prodigy.

BIOGRAPHY

Ines Montani is the lead developer of Prodigy, and a core contributor to spaCy. Although a full-stack developer, Ines has particular expertise in front-end development, having started building websites when she was 11. Before founding Explosion AI, she was a freelance developer and strategist, using her four years executive experience in ad sales and digital marketing.

15:00 — ALYONA MEDELYAN / THEMATIC

Fireside chat with John Henderson (Airtree Ventures)

BIOGRAPHIES

Alyona Medelyan is the Co-founder and CEO of Thematic, a YCombinator company, that helps companies improve customer loyalty and retention. She holds a PhD in Natural Language Processing and co-authored over 20 research publications. She worked on commercial projects in this field with companies like Google and Cisco before starting Thematic which now has customers like Stripe, AirBnB, Vodafone, Greyhound and Air New Zealand.

John Henderson is a Partner at Airtree Ventures where he focuses on vertical applications of data analytics. He led the firm's investments in Hyper Anna, Thematic, Earth AI and Flurosat. John was previously the COO of Summly, an NLP-powered news summarisation company acquired by Yahoo in 2013 and has also held roles at Facebook and BCG. He also co-founded London.ai, which has become Europe's premier AI meetup.

15:30 — BREAK & LIGHTNING TALKS

Jack Elliott / Flamingo AI

Andy Chisholm / Quin.ai

Jon Patrick / Health Language Analytics

Luiz Pizzato / Accenture Liquid Studios

Xavier Holt / Sypht

Will Radford / Canva

16:15 — PANEL WITH JOHN HENDERSON / AIRTREE VENTURES

16:45 — SOCIALISING

17:30 — CLOSE

Thanks to the [Australasian Language Technology Association](#) for supporting our international visitors, and [Tyro FinTechHub](#) for hosting. Organisation of this event is a volunteer effort provided by [Sydney NLP Meetup](#) founders Adam Schuck, Ben Hachey and Will Radford, with help from Sharon Lu and John Henderson.