

ROZETTA JOINS BIG TECH AT TOP NLP GLOBAL CONFERENCE IN IRELAND

Interview with Jacob Parnell, Data Scientist
by David Myton

The auditorium was packed with international Big Tech experts – representing business, industry and research.

Jacob Parnell, the 26-year-old postgrad from the northwest suburbs of Sydney was revealing new insights relevant to the esoteric world of Natural Language Processing (NLP) – the core of his work with the Sydney-based RoZetta Technology company.

The place was Dublin, Ireland, and the occasion was the highly-prestigious annual conference of the Association of Computational Linguistics, which was being held “in person” after a two-year, pandemic-induced gap during which organisers resorted to virtual delivery.

Watched by people from the likes of Google, Facebook, Apple, Amazon, Bloomberg, Spotify, and Grammarly, Jacob - expounding from a paper co-authored with RoZetta colleagues Inigo Jauregi and Professor Massimo Piccardi – presented new insights about a subject he’s come to know well: multi-document summarization.

The paper was titled A Multi-Document Coverage Reward for RELAXed Multi-Document summarization (MDS) – a process which aims to consolidate relevant points of information across a set of documents into an accurate, concise summary while maintaining readability.

Among other things, the paper spelled out the benefit of MDS for organisations working with large volumes of data:

- Reduced time to derive insights from text documents
- Capturing expert knowledge in a scalable system (Learns from the existing 'hand written' summaries)
- Consistency in the summaries generated - no human error
- Ability to ingest more sources without any impact to productivity
- More efficient use of high value resources - provides staff with more time to work on higher value opportunities.

That the paper was selected for presentation at the conference was a remarkable accomplishment. Out of 3,378 submissions, only 701 were accepted - and of those, just a select few were invited to give in-person talks.

Alongside working with RoZetta, Jacob is engaged in a PhD program at the University of Technology Sydney under the guidance of Professor Piccardi.

“For me it was the first conference I've actually attended in person for my PhD. It was just a great experience,” says Jacob.

“It was a huge event with thousands of people attending, either in person or virtually. There were several different streams with maybe hundreds of people attending the stream of their choice. And there were lots of industry partners there too, and it was great to chat with them.

“It’s easier to engage with someone face-to-face, talking about research over coffee or lunch. That’s just one aspect of the appeal of going to these conferences.”

With a background in astrophysics, maths and statistics, Jacob made the switch to researching NLP because he says,

“it’s where the world’s going in terms of machine learning and artificial intelligence”.

“I felt that the switch gave me the opportunity to cement my name as a researcher in a booming field.”

Essentially, NLP technologies - including computational linguistics and machine learning - allow computers to “understand” text and spoken words, complete with the speaker or writer’s intended sense and viewpoint.

“To give an idea of where NLP is at, you just have to look as far as Apple’s Siri and Amazon’s Alexa, and things such as online chatbot systems. Those are the prime examples of NLP applications.”

Jacob’s conference presentation looked at how new methods in “reinforcement learning” – training machine learning models to make a sequence of decisions - may improve the information extracted, or covered across a set of documents, during the summarization process.

With advances in NLP in recent years, machine understanding of text has dramatically improved as has the ability to automatically summarise content from individual documents or other text sources.

Continue to find out how RoZetta can help your business uncover value from your data in the second part of our interview with Jacob Parnell, Data Scientist at RoZetta Technology.

The paper outlined a new approach to the “tuning” of MDS models to materially improve the summaries generated in respect to their information coverage, which likely aims to suggest improvements in qualitative components such as consistency and relevance.

After his presentation, says Jacob, many of the conference attendees - including leading experts in the field - expressed interest in his research and also in the role of RoZetta Technology.

“Given that many people are familiar with the bigger companies such as Google Research, Apple, and the like, the visibility of a smaller company like RoZetta, appeared to be refreshing for many”

“Most were curious about the work that RoZetta does and how summarization is linked to this and plays a key application in its services.”

The conference, says Jacob, was a huge learning experience that he will be able to take into his work with RoZetta.

“It wasn’t just a talk-fest, it was a sowing and seeding of ideas – in fact, it was a melting pot of ideas.”

All the talks, presentations and papers, is a kind of a mixture of complete innovation and building on top of pre-existing ideas; to improve on what people originally thought couldn't be improved upon.”

An example of this, he says, was in the field of Information Retrieval.

“One paper I came across spoke to the point of improving the efficiency and speed at which you can extract information from documents - and that’s very useful from an industry perspective, because speed, efficiency and accuracy are absolutely everything.

“A topic like that is obviously useful and practical - it’s something that industry partners and so on are keen to listen to because they’ve got problems like that and want to solve them.”

HOW ROZETTA TECHNOLOGY HELPS BUSINESSES FIND THEIR BURIED TREASURE

It's a tough old world out there right now for most commercial and industrial enterprises as they endeavour to do business against a backdrop of elevated interest rates, soaring costs, and continued market uncertainty.

There's really nothing much you can do about it, but you can help to put yourself in the best position possible by ensuring you can access and utilise one of your greatest assets.

And that's the hard-earned knowledge and wisdom that's accumulated over the years and is sitting there unused in your archives.

Sydney-based RoZetta Technology's Jacob Parnell explains that the company employs leading edge technology to enable customers to access the hidden value that lies within the data contained in those archives.

"What really excites me about working for RoZetta is that we provide tangible solutions that really help our customers," says Jacob, whose expertise was on show recently when he was one of a select few invited to deliver a paper at the prestigious Association of Computational Linguistics conference in Dublin, Ireland, which he co-authored with RoZetta colleagues Professor Massimo Piccardi and Inigo Jauregi Unanue. (See article above)

RoZetta derives these solutions through the use of Natural Language Processing (NLP) techniques such as Multi-Document summarization, which has the ability to summarise the content within and across documents such as news releases, expert interviews, minutes, corporate announcements and press releases.



A key challenge for investors, businesses and analysts has been how to draw insights from this vast volume of information in a timely manner.



Jacob Parnell

"...now you can distill all of the noise that's around the key concepts that are conveyed in, say a transcript, and summarise 20 pages worth of documents into a nice half a page summary of what the content was about"

"...that's why I think RoZetta is in a very good place in that we're able to pinpoint certain problems that companies have in accessing their data, and try to help them by applying NLP."

Jacob says many industries "have a lot of textual data that they've just got sitting in their repositories that have heaps of information that is not necessarily utilised".

But with particular NLP techniques *"it becomes easier to extract the important bits from structured and unstructured data, to help drive the discovery of new information"*.

He adds, "Going back to the summarization example, it becomes less intensive for the user as someone doesn't have to go and read pages and pages of content to understand what the core message is...and that core message could be the very thing to help you navigate through these stormy times."

For inquiries contact us here or via email at enquiries@rozettatechnology.com

See separate story for more on the Multi-Document summarization (MDS) paper presented at the ACL 2022 Conference.