



DIGGING OUT THE GOLD BURIED IN YOUR DATA MINES

Interview with Scott Matthews,
Chief Data Scientist and Professor Massimo
Piccardi by David Myton

The American physicist William Pollard said that information is of little benefit “unless it is organized and available to the right people in a format for decision making”.

With that in mind, scientists and researchers at Sydney-based RoZetta Technology are helping businesses and organisations to reap the benefits of information stored in their own archives and repositories – but which to all intents and purposes has hitherto been inaccessible.

Often, this information lies in “unstructured” text – the likes of newspaper clips, press releases, corporate announcements, records of interviews, and so on.

Potentially there could be hundreds, perhaps thousands, of such documents in the archives. And they could – and often do – contain vital intelligence to inform decision making and market positioning.

“The problem is that you need someone to go in and read those documents to extract the information - and that is time consuming and expensive,” says RoZetta Technology’s Chief Data Scientist Scott Matthews.

“And even if you can get someone to read them all, that’s not necessarily going to produce consistent results.”

RoZetta Technology, explains Matthews, employs technology and data science centered on Natural Language Processing (NLP) and Software as a Service (SaaS) – to extract real and relevant meaning from potentially multitudes of varying sources.

“What we do is frame real-world problems in a mathematical context,” he adds.

“When you’ve got the documents into some sort of digital format, extracting the language from them is a fairly well-solved problem.”

RoZetta’s SaaS tool is able to unearth value from unstructured data by identifying what’s relevant and valuable and then linking those with structured data assets.

“Once unstructured and structured text are mapped, we are able to generate document summaries based on the document content,” says Matthews.

For business and industry, information gleaned from unstructured data – now conveniently identified and linked to existing structured data assets - can help to maximize value and provide a competitive advantage.

RoZetta Technology has worked with a number of companies to help to derive value from their banks of “unstructured data”.

“We have a long history of helping our customers by utilising technology,” says Matthews.

“We partner with them until we have resolved their issues with bespoke solutions crafted by our pool of data science experts.”

In one example, RoZetta worked with a leading research company that wanted to organise and understand masses of unstructured data – information from various sources including conference proceedings and media, and contained in varying formats – to boost its commercial activities.

Working alongside the company to understand its needs, the RoZetta’s data science team developed a customised Named Entity Recognition (NER) solution - a natural language processing technique that automatically scans entire articles and extracts predefined named categories of entities such as Organisations, Quantities, Monetary values, People’s names and company names.

“A research firm may have large volumes of analysis and expert opinions stored in various formats – and by applying data science techniques we can extract and gather key information to present relevant text from a group of documents to save time and increase the value offered,” says Matthews.

In another example, a team of RoZetta researchers worked with an SMS message provider to help it to understand how customers used its messaging platform.

Employing RoZetta’s Software as a Service (SaaS) platform DataHex, researchers sifted through millions of messages, which were systematically curated, classified and analysed

- allowing the provider to understand how customers use the messaging platform and thus boosting their business.

“We take a very consultative approach with our customers,” says Matthews.

“Working with them we ask what problems are we trying to solve, what is the research need, what’s the best solution – it’s always a collaboration.”

RoZetta Technology is especially well placed to stay on top of advances in relevant technologies thanks to its relationship with leading researchers in the field at the University of Technology Sydney (UTS) – in particular, Professor Massimo Piccardi, an expert in Natural Language Processing in the university’s Faculty of Engineering and IT.

Professor Piccardi works closely with the RoZetta data science team and helps to guide its research and development. He also supervises a number of postgraduate students employed by RoZetta.

He is co-author of a new research paper with RoZetta’s Dr Inigo Jauregi and Jacob Parnell, focused on Multi-Document Summarisation, which will be presented at the leading conference in the field in Dublin this year.

Professor Piccardi points out that RoZetta doesn’t seek to compete with US tech giants such as Google and Amazon.

“In fact, RoZetta builds on their work and adds value to it,” he says.

For more information contact us today.

See our separate story for more details about the research and technology in the field [here](#).

The key to unlocking this often-overlooked document treasury comes in the form of NLP – Natural Language Processing.

©2022 RoZetta Technology. All rights reserved.

2 of 2