

For this edition's column, I would like to talk about how the intelligence process can be used for both research and also sending out a message. Analytics is not just about numbers alone.

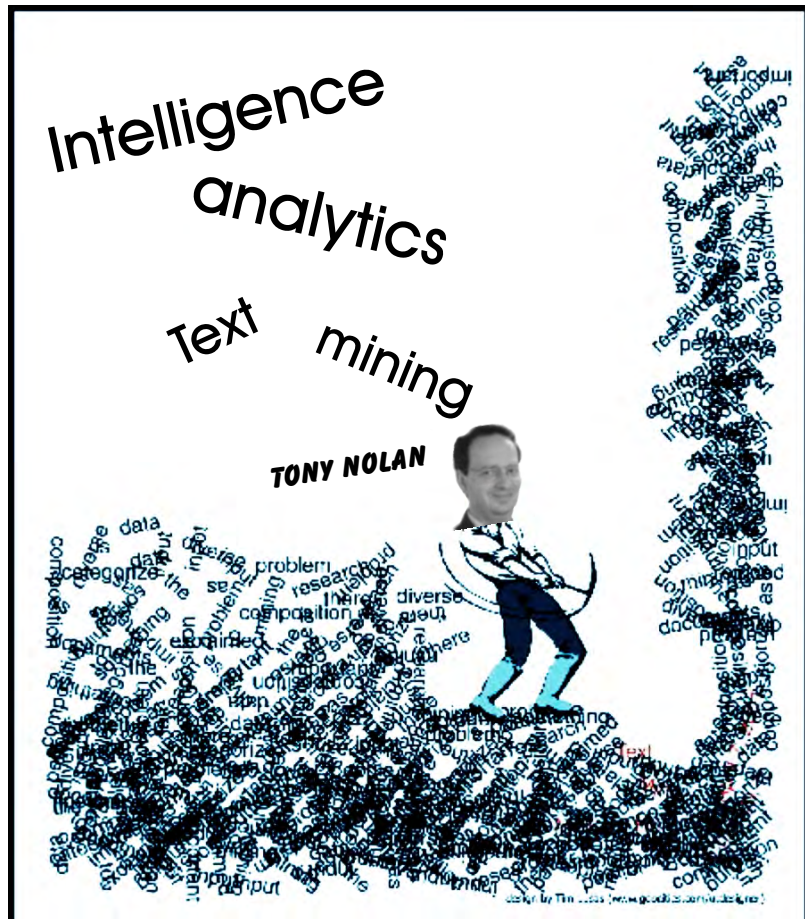
Text mining is the analytical process of understanding words, language, information and knowledge. While the majority of Intelligence work is about the words and language, very little analysis is conducted on text, as the equivalent of numbers, where you can look for patterns. The rest of this article is written on the assumption that you have legal authority and permission to conduct the following activities.

Text mining applies linguistic rules and statistical methods to automatically analyse electronic text and helps you discover insights that are often buried in large amounts of unstructured text data – such as email, social media content, blogs and other online media sources. **Text mining**, when combined with metadata (extra information about the dataset provided from outside of the dataset), gives a powerful view of the person's language, their intent, knowledge about the topic, and their intended message.

I bet you thought there is no room for human understanding in text mining, but there is. The Intelligence Officer is still needed to make sense of the high frequency words and concepts derived by nearest neighbour groupings to track the information behaviour of the targets.

You can text mine emails, text files, word documents, pdf's, and even webpages. There are computer programs like web spiders, web scrapers, etc, where you can put in the desired words your'e

Risk, Intelligence and Analytics Officer, Australian Government



interested in text mining, and they go off and use the search engines to locate those relevant webpages, and then save them in a corpus ready for text mining. There is also free text mining software such as **R** available.

Basically any one can use open source software for free, to locate a piece of text, or convert a pdf file to text, or download a webpage, and then process them into a dictionary, examine frequency counts, and look at the correlation between words, and have concept maps, link charts and word clouds. All you need is an internet connection, a Linux emulator, and the R console software. I will revisit R software in future articles.

Now for a real world example. The NSW Police Force in Australia has developed a project called **Eyewatch**. Basically, project Eyewatch is about communicating with the people of NSW using **Facebook** to gather Intelligence, and also to provide information back

to the general population, but more specifically the local population of each Local Area Command.

This type of project lends itself to text mining. Once the text is captured from both the Facebook posts and virtual neighbourhood watch meetings, then many different pieces of analysis are available. For instance once the data is loaded, you can compare sets of posts from different times, locations, sexes, communities, etc and examine if there are any highest frequency words, as well as outlier words, specific to each community.

You can also examine the language and the syntax between the datasets. Finally it also becomes possible to geocode the locations mentioned, and integrate it into your crime or marketing analysis. Of course you can also produce word clouds and link charts based on word frequency and nearest neighbour. •