

Natural Language Analysis Toolkit

Automated Text Processing

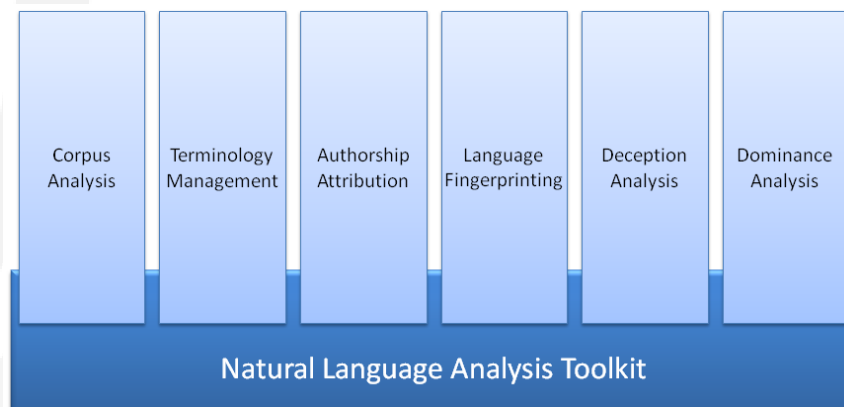


The Challenge

The Internet has grown at a phenomenal rate. Revolutionising how people communicate - for many this medium has now become the norm reflected in the rise of email, instant messengers and online social networks. A consequence is that a vast amount of text is being sent across the globe at any given time and this poses significant challenges when it comes to analysing such text for purposes such as policing or children's online safety. In many situations it is now unfeasible to manually analyse texts, whether it be due to limited available resources (for example, such as within law enforcement) or just due to the vast amounts of noisy data that is being dealt with. What is required, therefore, is a sophisticated automated approach that is able to complement and help to support the manual analysis process or, in some cases, replace it all together.

The Solution

Isis Forensics has developed a Natural Language Analysis (NLA) Toolkit that provides cutting edge analysis capabilities based on three decades of academic research. This toolkit uses hybrid statistical and rule-based methods to identify trends and to make comparisons that can achieve high degrees of accuracy.



The NLA toolkit has been developed in such a way that its various analysis functionalities can be configured and combined to provide a range of capabilities of various levels of sophistication. Specific scenarios include:

Corpus Analysis – used to analyse any spoken or written text to determine properties such as the most significant keywords, significant phrases, people and place names, semantic meanings.

Terminology lists generation and management – the ability to create and maintain terminology lists for a given topic based on the analysis of texts (for example, terminology used by paedophiles to describe child abuse activities). In particular the technology is able to cope with evolving terminology within a given domain, and intelligently identify new terminology that crops up.

Authorship attribution – analysing a text to statistically determine the characteristics of the author. This can include determining the gender, age group, country (i.e., British English, American English, etc), region and social class all in an automated fashion.

Language fingerprint of individuals and comparisons – extends authorship attribution further by developing a fingerprint of the author's language style. These fingerprints can be used to make comparisons, for example to statistically determine if two texts are written by the same person.

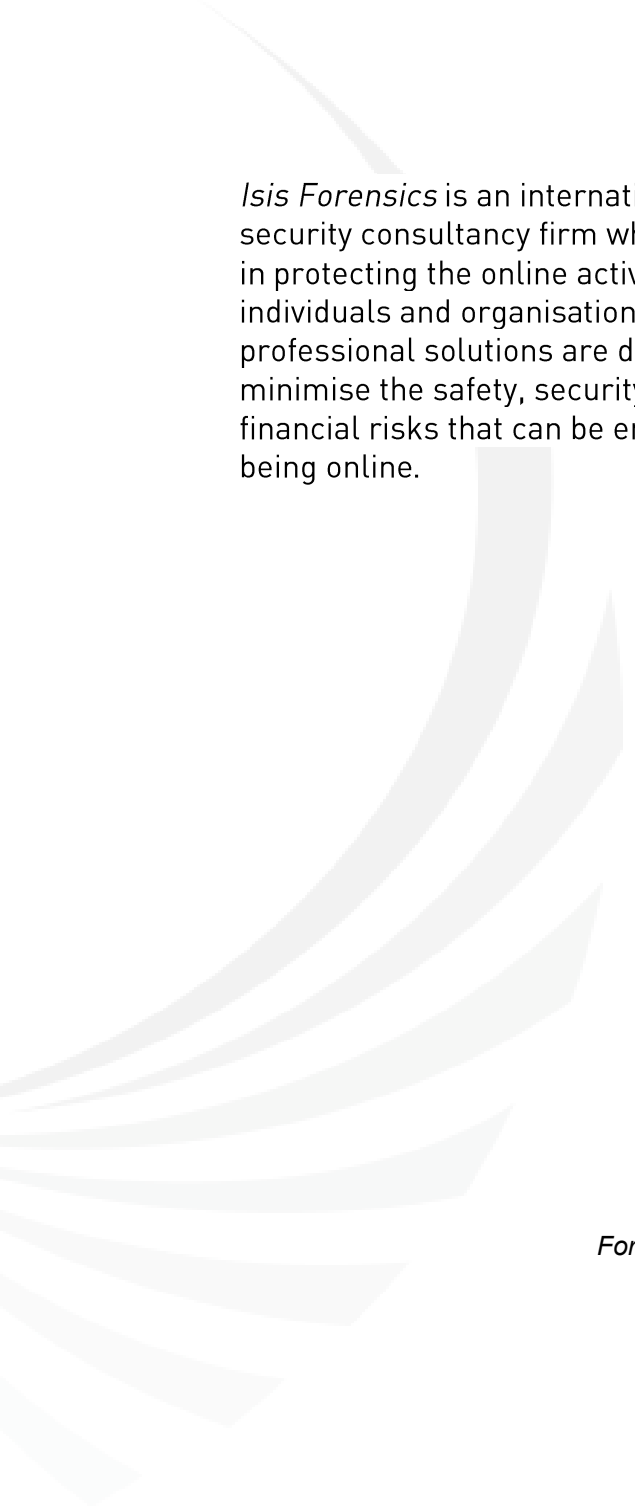
Deception analysis – analysing a chat log to statistically determine if deception is occurring in some form. For example, if one party within a two-way conversation is deceiving the other.

Dominance analysis – analysing a chat log to statistically determine the relationship between the communicating parties, and in particular if one is dominating the other. Among others, factors such as aggressiveness and passivity can be used to inform this analysis

The Benefits

Isis Forensics' Natural Language Analysis Toolkit provides a powerful mechanism for performing language analysis activities. The NLA Toolkit provides a number of key benefits:

- **Sophisticated** – the NLA Toolkit provides a set of analysis options that can be used to in a variety of situations where the analysis of text is required. From simple analysis functionality that can be used, for example, to highlight names or terminology within a given text, through to more sophisticated functionality that can be used, for example, to determine if a paedophile is masquerading as a child within a chat room.
- **Mass Data Analysis** – the NLA Toolkit allows for the mass analysis of data, reducing or even removing the need for slower and more costly manual analysis.
- **Configurable** – the NLA Toolkit is a collection of language analysis functionalities that can be used and combined in different ways to suit the needs of the given application. This can range from simple keyword identification through to combining the various analysis functionalities to perform more complex language analysis.



Isis Forensics is an international online security consultancy firm which specialises in protecting the online activities of individuals and organisations. Isis Forensics professional solutions are designed to minimise the safety, security, legal, and financial risks that can be encountered from being online.

For more information about how we can help you, contact us at:

Isis Forensics
The Knowledge Business Centre
InfoLab21
Lancaster University
LA1 4WA

Tel:+44(0)7092 891873
Fax:+44(0)7092 891895
Email:enquiries@isis-forensics.com
Web: www.isis-forensics.com