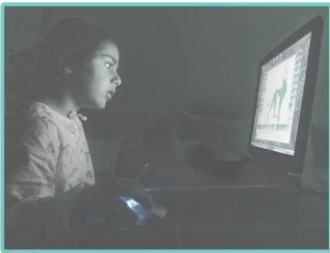


Language Forensics Toolkit



The Isis project has developed a forensics toolkit that can help law enforcement officers analyse the vast amounts of conversation data they collect as part of an ongoing investigation. The toolkit provides a range of powerful functionalities that can quickly summarise such data, extract intelligence and enable law enforcement personnel to focus on the high priority aspects of a case.

Key features of the toolkit include:

- **Data extraction and Management** – data can be imported from a variety of sources (IM, forums, email, etc). Conversations can also be captured in real-time (for example, during an under cover operation). Advanced techniques allow for the extraction of ‘chat fragments’ from disk and memory dumps, and their reconstruction into a conversation flow.
- **Chat log analysis** – sophisticated analysis allows investigators to quickly see who the participants are within a conversation, when the participants were online/offline, topics, names/places, etc, that are mentioned.
- **User Profile Building** – enables building profiles of potential suspects or victims. Sophisticated analysis of text determines age and gender characteristics with high accuracy, and extracts language stylistic data such as the key words or phrases a potential suspect or victim regularly uses.
- **Profile Comparison** – profiles can be compared against one another to identify potential similarities, for example, to help identify if multiple online personas are actually the same person
- **Timeline Analysis** – analyses data over a time scale, highlighting when specific users, names, places, terminology, etc occur. Side-by-side case analysis is also supported to help identify trends and common characteristics.
- **Terminology extraction** – automatically identifies new and evolving terminology for a given domain, so investigators are always aware of current language use.