


Please cite the Published Version

Paxton-Fear, Katie , Hodges, Duncan and Buckley, Oliver (2019) Increasing the accessibility of NLP techniques for Defence and Security using a web-based tool. In: Defence and Security Doctoral Symposium, 10 November 2019 - 11 November 2019, online.

Publisher: Cranfield University

Version: Presentation

Downloaded from: <https://e-space.mmu.ac.uk/627527/>

Additional Information: 2020 Defence and Security Doctoral Symposium (DSDS20) hosted by Cranfield University in conjunction with DSTL, AWE and FASS (Future Aviation Security Solutions)

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)



Increasing the Accessibility of NLP Techniques for Defence and Security using a Web-Based Tool

Katie Paxton-Fear K.Paxton-Fear@cranfield.ac.uk

What is Natural Language Processing (NLP)?

NLP techniques are the mechanisms by which a machine can process and analyse text written by humans. These methods are used for a range of tasks including categorising documents, translation and summarising text. To accomplish this a large volume of text (corpus) is collected, processed and finally models can be trained on the corpus. With current methods, the ability to manage corpora is rarely considered, instead relying on researchers and practitioners to do this manually in their file system. To train models, researchers write code directly for one task or experiment, limiting their reusability and ability to generalise.

The Problem: Low Accessibility

Although machine learning (ML) and natural language processing is becoming more common in Defence and Security, there are issues limiting its use. One such issue is the multidisciplinary nature of Defence and Security, with individuals from many backgrounds contributing to a single project. ML and NLP techniques require distinct specialisms to create and interpret a model. This is even more important when delivering research, where outputs may be operationalised and the accessibility can be a limiting factor in their deployment and use, limiting the impact of the work.

Solving the Problem

Managing Corpora

The corpora library allows a user to complete basic operations, including: creating a new corpus, changing the settings of an existing corpus and exploring the documents within a corpus. In addition it allows for complex topic modelling specific operations such as splitting a corpus into sentences and traversing a corpus.

Jargon

- Corpus/Corpora** – A body of documents used to create a model
- API** – A piece of software that is designed to communicate with software rather than being used by a human
- Training** – The process used to create machine learning models
- Insider Threats** – Security threats that arise from an organisations own employees rather than externally
- Topic Models** – An NLP model that attempts to automatically find topics in a piece of text using probabilities and key words

Creating Topic Models

Topic Models are models that attempt to split a piece of text into topics, this allows researchers to explore the content of text computationally. The topic model library allows for basic operations such as creating new topic models. However, users can also: import topic models that have already been created, traverse a topic model and delete a topic model. In addition models can be explored by topic and the sentences associated with each topic can be viewed.

Topic 49 (cookii privaci browser collect supra employe)

Document: 5c6e87b04128

The manager was involved in setting up the new IT system so that she could determine whether or not she could still continue her current fraudulent activities without being caught and when she realised this wasn't going to be the case she used her status and known personality to get around the new system and have her whole team exempt to avoid detection/questioning as to why only her and her accomplices required the exemption.

Document: 5c6e9bcbf644c

The higher management gave her department exemption to not use to system and enforced usage on the rest of the organisation.

Analysing Topic Models

Finally topic models can be applied to a corpus and analysed. *The Application Tool* allows a user to apply a model and view the results either:

- At a topic level, viewing sentences assigned a certain topic from different documents
- At a document level viewing how different sentences have been assigned different topics.

The Analysis Tool compares topic models by evaluating the symmetrical difference and the intersection, visualising how different models evaluate the same data.

Sentences

Corpus: perspectives

Model: event-extraction-150-topic-model

Intersection

These sentences appear in the topics selected

Appears With

She had done so over 18 years.

The news reports describe her being fined about \$63M, although there is no mention of a custodial sentence.

Not sure why, but it was mentioned that she had a lavish lifestyle and always bought drinks and meals for others.

The fraud include a further 9 accomplices, whose roles are still being investigated.

It would appear that she may have had other members of staff write the cheques and she could authorise them however this is not confirmed.

The manager was involved in setting up the new IT system so that she could determine whether or not she could still continue her current fraudulent activities without being caught and when she realised this wasn't going to be the case she used her status and known personality to get around the new system and have her whole team exempt to avoid detection/questioning as to why only her and her accomplices required the exemption.

Summary

- Using natural language processing (NLP) tools can be difficult for non-specialists.
- Defence and Security involves many different people with different backgrounds
- Machine learning is becoming more common within the defence and security domain
- Therefore it is necessary to create a tool which offers the same functionality as ad-hoc code but is clear and approachable without requiring specialist skills
- Solution: Web based tool with R application programming interface (API)**

Technology Stack

Front End – Bootstrap Framework
 Back End – PHP7 and the Laravel Framework
 NLP API – R using Plumber
 Database – PostgreSQL

Katie Paxton-Fear K.Paxton-Fear@cranfield.ac.uk †, Dr Duncan Hodges d.hodges@cranfield.ac.uk †,

Dr Oliver Buckley o.buckley@uea.ac.uk ‡

† Cranfield University ‡ University of East Anglia

Centre for Electronic Warfare and Cyber, Cranfield Defence and Security, Defence Academy of the United Kingdom

www.cranfield.ac.uk

