Identifying Changes in the Cybersecurity Threat Landscape using the LDA-Web Topic Modelling Data Search Engine

Thursday 13th July 2017

Multidisciplinary approaches to Cloud Crime

HCII 2017, Vancouver Canada

Noura Al Moubayed, David Wall, and A. Stephen McGough

Outline

- **The Problem**
- Text Processing for Topic Modelling
- Using Topic Modelling for searching

The Problem

- "90% of all the data in the world has been generated over the last two years"… IBM

- "85% of worldwide data is held in un-structured formats"…
Berry and Kogan

- How can we understand it? ….or better still make use of it?

- How can we determine the most pertinent information? …and then act on it?

- How can we find the needle if we are not sure what it looks like or what hay looks like?

- "Without labelling, you cannot train a machine with a new task"… IBM

Outline

- The Problem
- **Text Processing for Topic Modelling**
- Using Topic Modelling for searching

Topic Modelling: Latent Dirichlet Allocation (LDA)

Outline

- The Problem
- Text Processing for Topic Modelling
- **Using Topic Modelling for searching**

Demo –
Topic Modelling

Durham University

Search Engine based on Topic Modelling

Upload file

Choose File   no file selected

User Query

No. Displayed Results

100

Search

Return to Home Page

Potential Applications:

- Security Applications:
  - Terrorist activity tracking
    - Acting out of character, predicting activity
  - Police crime database
    - Criminal profiling, acting out of character
  - Unwanted information release
    - Topic changes, specific damaging subjects

- Other
  - Student applications
    - Identifying bogus attempts for visa
  - Social Media tracking
    - Social grooming, political persuasion, product complaints
  - Fake News identification and tracking
  - Sentiment tracking

  …..Your use-case here

We Are recruiting:
- 1 PostDoc (Machine Learning / NLP)
- 1 PostDoc (Parallel Programming)
- Always looking for good PhD Candidates

stephen.mcgough@newcastle.ac.uk
noura.al-moubayed@dur.ac.uk
D.S.Wall@leeds.ac.uk