



Defence Digital Service Centre of Expertise

CASE STUDY

**Methods
Analytics_**

THE CHALLENGE



The MOD holds an enormous amount of unstructured text data.

The MOD holds an enormous amount of unstructured text data. One of the barriers to using standard models to gain insights from this data is the extensive use of bespoke acronyms: there are more than 21k acronyms commonly in use across the MOD, with some of them representing up to 17 different meanings. This complexity means creating a straightforward acronym dictionary was impossible. They needed a tailored solution to recognise and retrieve the long explanation of acronyms in text.

21K+

There are more than 21k acronyms commonly in use across the MOD.

17

Some of them representing up to 17 different meanings.

TECHNICAL DETAILS

We used Python SpaCy models to identify entities and associate them to their long definitions via the Named Entity Linking techniques. When multiple matches were identified between an acronym and its definition, semantic vectorisation and semantic similarity were used to identify the match with highest probability. This technique transforms terms into numbers to look at how mathematically similar these are in a n-dimensional space (where they are represented as vectors).




THE APPROACH

Methods Analytics worked with data scientists at DE&S to understand the requirement and develop a solution on publicly available Defence documents. An object-oriented Python pipeline was then used to train an AI model that recognised acronyms and linked them to their long definition. In case a single acronym pointed at multiple definitions, the context in which it was used provided the correct match along with a reliability score.

THE OUTPUT

Methods Analytics delivered a pipeline that has been integrated in the MOD Natural Language Processing pipeline and is now available within for internal MOD network. We provided an interactive solution which delivered insights on long documents in a consistent and robust way. This saved MOD personnel precious time, allowing them to label and re-use documents quickly and to spend more time in tasks where human intervention is necessary. This project was presented to the 2021 AI Fest and a video can be found [here](#).

- ① Saved MOD personnel precious time
- ② Re-use documents quickly
- ③ More time in tasks where human intervention is necessary



We're Methods Analytics – a specialist data services company. We use data to help public and private sector clients solve complex problems and do good things.

**Methods
Analytics** 

© Methods Analytics Ltd.
Methods Analytics (Middle East) Ltd.
All rights reserved

Registered in England and Wales
Company No. 2485577
VAT No. GB 859 1403 14

Registered in Abu Dhabi
Company No. 000004100

analytics@methods.co.uk

METHODSANALYTICS.CO.UK