

DSC-72 Data science for NICE guidance

2018-08-15

This project targets the ongoing ‘surveillance’ of guidance recommendations through the following search functionality:

Given a recommendation, retrieve similar or related recommendations
Given a set of keywords, retrieve related recommendations
Given a set of keywords, retrieve a wider set of related keywords
A data driven classification of terms will also be investigated to assist in this search.

Team members

- Jasmine Latham
- Thanasis Anthopoulos
- Bernard Peat
- Ian Grimstead
- Rubin Henstra-Hill

The need

To automate the NICE guidance development and review process to: * decrease the length of time it takes, cutting costs * make updates more active than reactive * improve the quality and scope of intelligence gathering and search

Impact

NICE currently does not have a robust mechanism for identifying related products within the portfolio. At present linkage is judgement generally based on guideline title and high level headings rather than individual recommendations. This can lead to inconsistency for example in the guideline portfolio with overlapping recommendations in different products sometimes contradicting each other. Automated linkage should ultimately also allow for easier repackaging of content into different products. Further advantage may also be realised if a suitable evidence management database could utilise this linkage to reuse evidence.

Data science

- Web scraping (BeautifulSoup)
- Topic analysis (SciKit Learn: NMF, LDA)
- Word / sentence embeddings (HuggingFace: transformers - e.g. BERT, RoBERTa)
- Semantic text similarity (UKPLab: sentence transformers)

Stakeholders

- NICE - Kay Nolan, Fiona Glen
- Likely use case for OGDs

Code and outputs

- repo
- planning
- slack

Delivery

- [X] May 2019 Prioritised user stories agreed with stakeholder
- [X] June 2019 NICE syndication license signed off (data access)
- [X] June 2019 Project started
- [X] September 2019 data engineering completed
- [X] October 2019 recommendation V&V data provided by stakeholder
- [X] December 2020 first recommendation matches provided to stakeholder
- [X] January 2020 gold standard for V&V input recommendation matches
- [] March 2020 final model delivery for recommendation matching
- [] March 2020 keyword search support delivered to stakeholder

Further information

Please contact datasciencecampus@ons.gov.uk for more information.

Updates

- No updates yet.