Poster: An exploration of NLP and NER for enhanced search in osteoarchaeological and palaeopathological textual resources

Alphaeus G W Talks (University of York)\*

[AlfieTalks@live.co.uk\*](mailto:AlfieTalks@live.co.uk*)

Abstract:

Background:

The advent of Big Data represents an ever-increasing opportunity, particularly in the world of Archaeology. Archaeology is a destructive process; text-based reports are written to document the site and the resulting ‘grey literature’ contains a wealth of information for scholars and researchers. This information becomes part of data archives such as the Archaeology Data Service, but currently, there is no effective way to navigate the content of these reports, meaning that these databases become labyrinths containing huge amounts of information that are difficult to access. The technology used to navigate the Big Data within these archives is ever-improving in its accessibility. These improvements include the use of Natural Language Processing (NLP) to create enhanced metadata and these are mapped to controlled terms using Named Entity Recognition (NER) (Richards et al. 2011, 31-56). The work represented in this poster is the continuation of projects carried out by the Archaeology Data Service and Talboom (Meghini et al. 2017, 1–27; Talboom 2017).

Subject:

This poster will present the two technological approaches of NLP and NER to increase the accessibility of palaeopathology and osteoarchaeological data. These have been analysed for accuracy and usefulness by professional bioarchaeologists, students and laymen. From the results, despite some limitations, it will be shown that there is real potential in the use of NER and NLP to allow osteoarchaeology and palaeopathology information to be accessed more easily, thus unlocking the information trapped within Big Data. The results further indicate that the use of the technology will generally allow the average layman to better connect to the resources. This is verified by a quantitative survey of experts, students and laymen of osteoarchaeology and palaeopathology. The survey included a selection of five 7-point Likert scale questions, addressing factors such as time-saving, reliability, accessibility, usability and usefulness, Thus these technologies may hold the future of connecting archaeology to the public and encouraging further public engagement.

Discussion:

The results, therefore, show the implications of the application of NER and NLP to osteoarchaeological and palaeopathological records. They enable the records to be accessed by all, despite being held in previously inaccessible archives. It also reveals that NLP has the potential to become an incredibly useful tool to address the Big Data created by archaeological projects. If such methods were regularly implemented, research questions regarding specific details from many sites could be answered with greater speed and accuracy. These technologies will, therefore, be central to the future of archaeological research. CAA 2021 “Digital Crossroads”

References:

Meghini, Carlo, Franco Niccolucci, Achille Felicetti, Paola Ronzino, Federico Nurra, Christos Papatheodorou, Dimitris Gavrilis, Maria Theodoridou, Martin Doerr, Douglas Tudhope, Ceri Binding, Roberto Scopigno, Andreas Vlachidis, Julian Richards, Holly Wright, Guntram Geser, Sebastian Cuy, Johan Fihn, Bruno Fanini and Hella Hollander. 2017. “ARIADNE: A Research Infrastructure for Archaeology.” Journal on Computing and Cultural Heritage 10, no. 3: 1–27. <https://doi.org/10.1145/3064527>.

Richards, Julian, Stuart Jeffrey, Stewart Waller, Fabio Ciravegna, Sam Chapman and Ziqi Zhang. 2011. “The Archaeology Data Service and the Archaeotools Project: Faceted Classification and Natural Language Processing.” In Archaeology 2.0 new approaches to communication & collaboration, edited by Eric Kansa, Sarah Whitcher Kansa and Ethan Watrall, 31-56. California: Cotsen Digital Archaeology.

Talboom, Leontien. 2017. “Improving the discoverability of zooarchaeological data with the help of Natural Language Processing.” Archaeological Information Systems, University of York.