From text to data: a NLP approach to digital archaeology.

Elisa Paperini, Nevio Dubbini, Gabriele Gattiglia

Extracting data from archaeological texts (from grey literature to journal papers) represents one of the archaeology's most leading challenges. In recent years, Natural Language Processing (NLP) has been also adopted in the archaeological domain, but we are still far away from achieving robust results. This work is part of a more complex project related to extraction, visualisation and analysis of text data, carried out by MAPPA Lab, a digital archaeology lab of the University of Pisa, together with Italian NLP Laboratory of the Institute for Computational Linguistics "A. Zampolli" (http://www.italianlp.it/). The aim of this work is to set up an as-automatic-as-possible procedure to overcome one of the main barriers to data accessibility, related to digitisation of data in a way allowing to process them. We developed a semi-automated workflow for text extraction and processing of data from pdf to a previously designed RDBMS. The extraction of data about location, date, authors, bibliography, archaeological findings and chronology was texted on about 1300 short communication papers (about 120000 text lines) published in the Italian journal of Medieval Archaeology (Archeologia Medievale), from 1974 to 2017. A formalised vocabulary of archaeological terms was first developed, then text extraction and NLP algorithms were applied, in order to detect, tag and insert the extracted data into the database. This method of working on data can be applied to all sources on which it is necessary to carry out similar research activities. Moreover, retrieved data are digital, accessible and reusable.

Surveying the Okak Inuit Winter Village using Exploratory Photogrammetry

James Samuel Williamson, Francois P. Levasseur, Peter Whitridge

In July and August 2018, a team from Memorial University of Newfoundland surveyed the prominent Inuit winter village site of Kivalehk, near Okak in northern Labrador. UAVs and an Emlid RTK were used to digitally record the presence of numerous burials, caches and semi-subterranean house features throughout the site to analyse their morphology and distribution, and to evaluate the efficacy of differing methods of photogrammetry, which will be discussed in this presentation. The first method involves a near real-time survey technique which allows the surveyor to target areas and features using photogrammetry, GIS, and handheld GPS tools for subsequent high-resolution photogrammetric recording and ground-level inspection. The second method tested the most useful means of data acquisition for the purposes of image-based modelling and to put forward a workflow for video-based capture using 4K cameras. The outcome of this research will lead to the possibility of improving photogrammetric modelling of landscapes, and to the ability to target landscape phenomena using consumer grade UAVs and RTKs.