Digging by Debating Final Report Abstract

The Digging by Debating (DbyD) project aims to identify, extract, model, map and visualise philosophical arguments in very large text repositories such as the Hathi Trust. The project has: 1) developed a method for visualizing points of contact between philosophy and the sciences; 2) used topic modeling to identify the volumes, and pages within those volumes, which are 'rich' in a chosen topic; 3) used a semi-formal discourse analysis technique to manually identify key arguments in the selected pages; 4) used the OVA argument mapping tool to represent and map the key identified arguments and provide a framework for comparative analysis; 5) devised and used a novel analysis framework applied to the mapped arguments covering role, content and source of propositions, and the importance, context and meaning of arguments; 6) created a prototype tool for identifying propositions, using naive Bayes classifiers, and for identifying argument structure in chosen texts, using propositional similarity; 7) created tools to apply topic modeling to tasks of rating similarity of papers in the PhilPapers repository. The methods from 1 to 5 above, have enabled us to locate and extract the key arguments from each text. It is significant that, in applying the methods, a non-expert with limited or no domain knowledge of philosophy has both identified the volumes of interest from a key 'Big Data Set' (Hathi Trust) AND identified key arguments within these texts. This provided several key insights about the nature and form of arguments in historical texts, and is a proof-of-concept design for a tool that will be usable by scholars. They have further created a dataset with which to train and test prototype tools for both proposition and argument extraction. Though at an early stage, these preliminary results are promising given the complexity of the task.

Specifically, we have prototyped a set of tools and methods that allow scholars to move between macro-scale, global views of the distributions of philosophical themes in such repositories, and micro-scale analyses of the arguments appearing on specific pages in texts belonging to the repository. Our approach spans bibliographic analysis, science mapping, and LDA topic modeling conducted at Indiana University and machine-assisted argument markup into Argument Interchange Format (AIF) using the OVA (Online Visualization of Argument) tool from the University of Dundee, where the latter has been used to analyse and represent arguments by the team based at the University of East London. This work has been articulated as a proof of concept tool – linked to the repository PhilPapers – designed by members linked to the University of London. This project is showing for the first time how big data text processing techniques can be combined with deep structural analysis to provide researchers and students with navigation and interaction tools for engaging with the large and rich resources provided by datasets such as the Hathi Trust and PhilPapers. Ultimately our efforts show how the computational humanities can bridge the gulf between the "big data" perspective of first-generation digital humanities and the close readings of text that are the "bread and butter" of more traditional scholarship in the humanities.