**Technical Appendix**

All files, including an interactive mapmaker, are available along with more detailed background and instructions for replication in a Github repository: <https://github.com/mapping-british-empire/MappingBritishEmpire>.

*Corpora*

The *CLIWOC* and *Slave Voyages* data were downloaded from their respective websites. Text files of the texts were pulled from either Project Gutenberg, Gale’s Digital Scholars Lab (DSL), or the National Library of Australia’s South Seas collection of Pacific travel writing (table 7).

Table 7: Sources for Digitized Texts

|  |  |  |
| --- | --- | --- |
| Gutenberg | DSL | South Seas |
| Daniel Defoe, *Robinson Crusoe* (1719); Defoe, *Further Adventures of Robinson Crusoe* (1719); William Rufus Chetwood, *Richard Falconer* (1719); Defoe, *Captain Singleton* (1720); Captain Charles Johnson (?), *A General History of the Pyrates* (1724); Defoe, *A New Voyage Round the World* (1725); Jonathan Swift, *Travels into Several Remote Nations of the World (Gulliver’s Travels*) (1726); George Anson (edited by Richard Walter), *A Voyage Round the World* (1748);  Tobias Smollett, *Roderick Random* (1748) | William Rufus Chetwood, *Richard Falconer* (1719); John Green (editor), *New General Collection of Voyages* (1744); Tobias Smollett (editor) (?), *A Compendium of Authentic and Entertaining Voyages* (1756); Abbé Raynal and John Justamond (translator), *A Philosophical and Political History…of the Europeans in the East and West Indies* (1776) | Joseph Banks, journal of *Endeavour* voyage and description of places (1768-71); James Cook, journal of *Endeavour* voyage and description of places (1768-71); John Hawkesworth, *An Account of the Voyages for Making Discoveries in the Southern Hemisphere* (1773); Sydney Parkinson*, A Journal of a Voyage to the South Seas* (1773) |

Texts from the DSL were produced using uncorrected optical character recognition (OCR) and were scrubbed using the Stanford Literary Lab’s version of Ted Underwood’s code. References to novels within the body of the essay use the page numbers of the editions in the bibliography.

*Mapping*

Maps were made in R using the following packages: data.table; ggmap; leaflegend; leaflet; maps; mapdata; maptools; marmap; raster; rgdal; lattice; shiny. Remi Daigle’s Github page provided the skeleton for the program, which was then put into an interactive Shiny program. Locations were pulled from *CLIWOC* and *Slave Voyages* directly, and from the texts using Stanford NER as explained below. Specific land locations (cities, islands, capes, small sub-national regions), specific sea locations (bays, straits, rivers), and vague locations of both types (nations, empires, continents, the East and West Indies, oceans, seas) were differentiated according to my judgment of what might be specific enough to be navigationally useful to a mariner. Locations were manually associated with latitudes and longitudes in a master list, with vague locations roughly centered on the place in question.

*Place Names*

Place names were extracted from the texts using a custom Stanford NER model trained on a selective subset of the sea fiction corpus, as explained in the Stanford NER CRF FAQ. Note that this new “place name” entity was trained from a blank start rather than upon any of Stanford NER’s existing classes. In the performance metrics below, the recall is the most important figure, as I reviewed the tagged place names by hand to assign them a geographic location. The performance metrics for my custom model (table 8) are significantly better than the default Stanford NER and slightly better than Wilkens’s on his corpus of American fiction from 1851-75, and are good results given that human annotators only agree on identifying named locations 80 to 90% of the time (Leidner 10, quoted in Wilkens 840).

Table 8: Performance Metrics for Identifying Place Names

|  |  |  |  |
| --- | --- | --- | --- |
|  | Recall | Precision | F-score |
| Default Stanford NER | .48 | .83 | .61 |
| Trained Stanford NER | .69 | .83 | .75 |
| Wilkens (Built-in Stanford NER Applied to 1851-1875 US Fiction) | .65 | .81 | .72 |