Mapping the Journeys of Australian World War I Soldiers using Named Entity Recognition

Ashley Dennis-Henderson, Matthew Roughan, Jonathan Tuke

Over 330,000 Australians served abroad during World War I (Gammage 1990: 313). Generally, soldiers began their service with basic training in Australia, before travelling overseas for further training (DVA Anzac Portal 2023). Early in the war, Australians were sent to Egypt for this training. Then, in 1915 the Australian and New Zealand Army Corps (ANZACs) began the Gallipoli campaign, where, working from Lemnos Island as a base, they launched an attack on the Turkish shores. At the conclusion of this campaign, Australians travelled back to Egypt for further training before serving either on the Western Front or in the Middle East. Later in the war, Australian soldiers were also trained in England before serving on the Western Front. Likewise, the Australian Flying Corps primarily trained in England.

Whilst all Australian soldiers in World War I travelled along a similar route, each soldier had a personalised journey. Soldiers could be injured, taken prisoner, or killed. Furthermore, soldiers were given leave, and could be assigned to a wide variety of units and roles. These individual journeys are highlighted in their war diaries. During the war many soldiers kept diaries of their experiences, including details about their movements. Through a close examination of the diaries, it would be possible to map each soldier's journey through the war. The aim of this work is to do this automatically using named entity recognition (NER).

The diaries used in this analysis come from the State Library of New South Wales (SLNSW) collection. After the war finished, the SLNSW began collecting documents related to the service of Australians during the war (Edmonds 2020: 187). This collection contains 504 diaries written by 195 individuals, which have been digitised, transcribed, and cleaned, allowing digital analysis methods to be performed.

In this analysis, locations are initially extracted using named entity recognition. The locations are then cleaned and geocoded. By linking the extracted locations with the dates they were written on, the journey of each soldier can then be mapped over time. There are several important factors that need to be considered when cleaning and geocoding the extracted locations. Firstly, whilst literacy levels in Australia were greatly improved by the time World War I started, many of the diaries include spelling errors. Furthermore, many soldiers shortened words within their diaries, for example, writing "Australia" as "Aus". As such, a necessary cleaning step is matching the various forms of the same location and correcting their spelling. Another important cleaning step is to remove generic locations, such as "town", and non-locations that are extracted by the NER. After this is complete, it is necessary to determine which of the remaining locations were locations that the soldier visited, as opposed to the soldier talking about events happening elsewhere, and correctly geocode these locations. Difficulties arise in geocoding the locations as many places around the world share the same name. For example, there is a Liverpool in Australia and in the UK, both of which could have been visited by Australian soldiers. Both of these issues can be addressed by combining the date the location was written with information about the surrounding text. For example, soldiers could only travel so far per day. Therefore, if on one day a soldier mentions both Cairo (Egypt) and Melbourne (Australia), it is not possible that the soldier visited both those locations on the same day. Furthermore, if the locations mentioned on the days surrounding this date are all from Egypt, this would suggest that the soldier did not visit Melbourne on the date in question.

References

Gammage, Bill (1990): The Broken Years: Australian Soldiers in the Great War, Illustrated ed. Ringwood: Penguin Books.

DVA (Department of Veterans' Affairs) Anzac Portal (2023): "Training Australian army recruits during World War I." https://anzacportal.dva.gov.au/wars-and-missions/ww1/military-organisation/training. Accessed 08.12.2023.

Edmonds, Elise 2020. "Of sentimental value: collecting personal diaries from the First World War. Archives and Manuscripts 48, no. 2, 186-199. https://doi.org/10.1080/01576895.2020.1755701.

Picture Credit

j4p4n (2021): *Aussie World War I Soldiers Silhouette*, https://openclipart.org/detail/331587/aussie-world-war-1-soliders-silhouette. Accessed 24.07.2024.