

# Graduate School of Science and Technology Master's Thesis Abstract

Laboratory name (Supervisor)	Natural Language Processing (WATANABE TARO (Professor ))		
Student ID	2211410	Submission date	2024 / 1 / 18
Name	EUNIKE ANDRIANI KARDINATA		
Thesis title	Constructing an Indonesian-English Travelogue		
Abstract			
<p>Research in low-resource language are often hampered due to the under-representation of how the language is being used in reality. This is particularly true for Indonesian language because there is a limited variety of textual dataset and majority were acquired from official sources with formal writing style. All the more for the task of geoparsing, which could be implemented for navigation and travel planning applications, such dataset are rare, even in the high-resource language such as English. Being aware of the need for a new resource in both languages for this specific task, we constructed a new dataset consisting both Indonesian and English from personal travelogue articles. Our dataset consists of 88 articles, exactly half of them written in each language. We covered both named and nominal expressions of four entity types related to travel: location, facility, transportation, and line. We also conducted experiments by training a classifier to recognise named entities and their nominal expressions. The results of our experiments showed a promising future use of our dataset as we obtained F1-score above 0.9 for both languages.</p>			