1 Introduction

As the amount of published articles and other Dutch text on the Internet continues to increase, it would be helpful for researchers and others interested in location-specific articles to be able to search for articles based on locations mentioned in the article. It would enable them to spend less time sifting through irrelevant articles, and to spend more time doing their own research. A geolocation model could also be beneficial for other purposes, such as automatic traffic accident location statistics.

? have attempted to accommodate this need by creating an application, LocLinkVis, that plots Dutch input text on a map based on geographical entities in the text. By doing so, users can more quickly see if an article is relevant. The approach taken by **?** identifies toponyms in text by using a gazetteer populated with OpenStreetMap data. After toponym disambiguation, the most likely candidate is shown on a map. However, LocLinkVis does not take into account that some locations might consist of more than one toponym. In such a case, the application simply shows the location of all involved toponyms, instead of understanding how these toponyms are linked and need to be interpreted as one location.

1.1 Research question

This research is aimed at improving the accuracy of geolocating complex locations mentioned in Dutch articles, through the use of toponyms and surrounding terms containing spatial information. As explained in **?**, toponyms are most easily described as place names. The following excerpt is a definition set by the United Nations: "*A geographical name may also be referred to as a topographical name or toponym*", (**?**). The surrounding terms containing spatial references about toponyms are mostly prepositions of place, but can also be other adjectives or nouns. In this research, this group of terms will be referred to as *spatial adpositions*.

Accurately geolocating complex locations is especially important, as non-complex locations are already handled with sufficient accuracy by commonly-used geographical search engines such as OpenStreetMap and Google Maps. In this research, complex locations are defined as locations that involve multiple toponyms to accurately describe. Toponyms can refer to various geographical levels, such as streets, cities or regions. What makes a location complex, is the fact that there is usually one geographical scope to which more than one toponym refers, thus making it difficult for a model. To understand how these toponyms relate to each other, the previously defined spatial adpositions have to be analyzed. After analysis, it should become clear how all toponym coordinates can be combined to obtain the complex location. Below is an example of a complex location:

"De Burgemeester Bechtweg is tussen de Nederlandweg en de aansluiting met de A65 in beide richtingen afgesloten."

In the example above, the location is not the entire length of the road, but only between two other toponyms. Such locations are usually not accurately geolocated by the common geographical search engines. This research thus aims to answer the following question:

Using spatial adpositions to model the relationship between toponyms, how accurately can complex locations be geolocated?

1.2 Overview

[NOG SCHRIJVEN]