

# Transformer Based Geocoding

Yuval Solaz\*  
yuval.solaz@gmail.com

Vitaly Shalumov\*  
vitaly.shalumov@gmail.com

October 30, 2022

## Abstract

In this paper, we formulate the problem of predicting a geolocation from free text as a sequence-to-sequence problem. Using this formulation, we obtain a geocoding model by training a T5 encoder-decoder transformer model using free text as an input and geolocation as an output. The geocoding model was trained on geo-tagged wikidump data with adaptive cell partitioning for the geolocation representation. All of the code including Rest-based application, dataset and model checkpoints used in this work are publicly available.

**Keywords**— language model, sequence-to-sequence, transformer, geolocation, geocoding

---

\*Equal Contribution