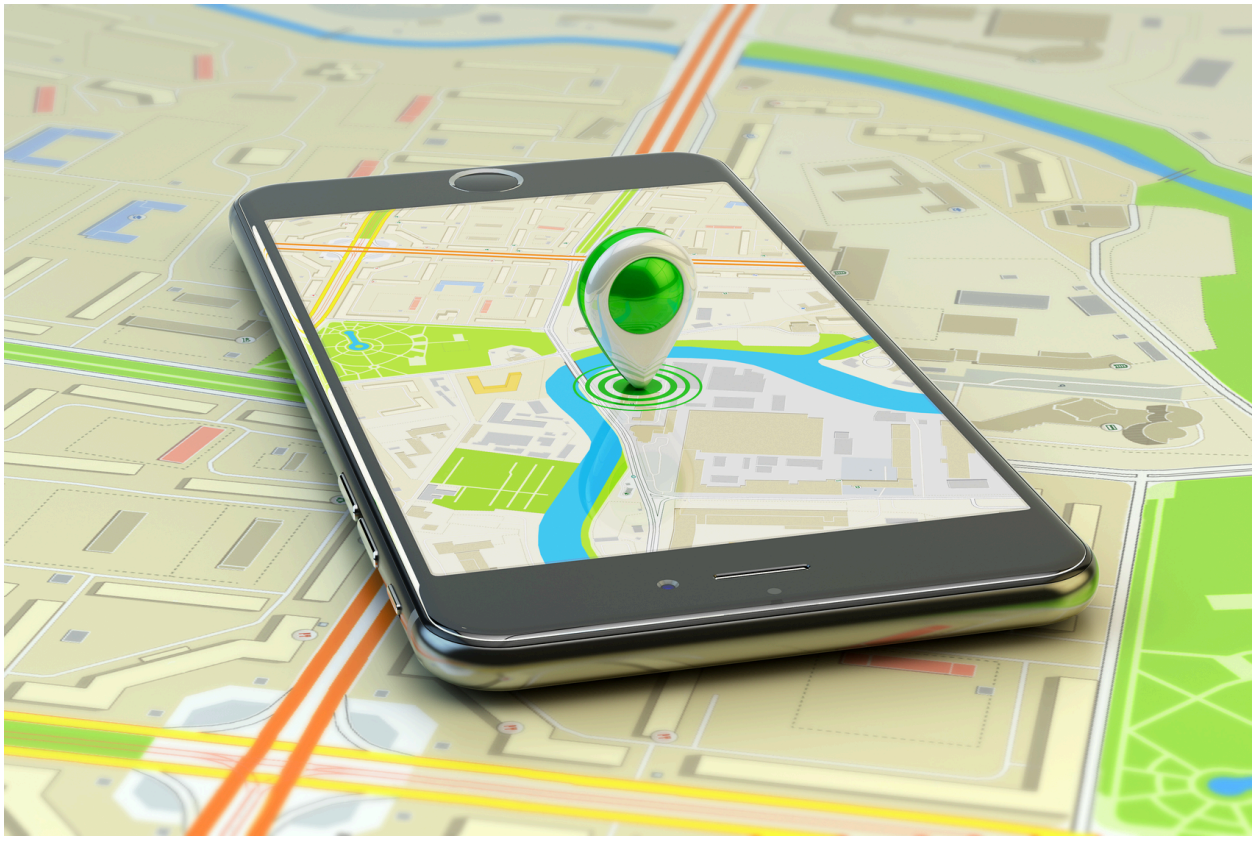
GEOINSIGHT: Unveiling Location Clues for Enhanced Intelligence

University of Auckland | SIGNAL TEAM 7



Source: JournalsOfIndia

INTRODUCTION

CLAVIN-NERD (CLAVIN Named Entity Recognition and Disambiguation) is a geotagging and geoparsing system focusing on Named Entity Recognition (NER). It aims to identify location-related entities from text and disambiguates them to specific locations mentioned in the reference text.

This report suggests some methods of improvement of the CLAVIN-NERD model, focusing on the following features:

Project Outcomes	Key Activities to Achieve the Outcome
Text Recognition Improvement	1. Propose multiple input text handling methods to support better correlation between use cases. 2. Improve atypical text processing with updated procedures.
Environment Recognition Improvement	1. Improve address processing methods to enable extraction of specific street addresses from a given input. 2. Recognise and map textual references to points of interest accurately. 3. Refine forms of identification to better distinguish between ambiguous locations and similarly named entities.
Location Support Expansion	1. Endorse dataset expansion to recognise more non-euro-centric locations. 2. Enhance the model to handle the correlation of multiple input texts for better location support expansion.

TEXT RECOGNITION IMPROVEMENT

Text.

ENVIRONMENT RECOGNITION IMPROVEMENT

Text.

LOCATION SUPPORT EXPANSION

The employment of location support expansion techniques is essential to enhance the accuracy and performance of CLAVIN-NERD. While conventional methods can identify basic locations, they may fail to capture diverse location types.

A few suggested methods to improve the performance of CLAVIN-NERD for location support expansion are:

1. ***Additional Gazetteers and Knowledge Graphs:*** Gazetteers and knowledge graphs offer significant information about locations, including hierarchical relationships, alternative names and geographical coordinates. Integrating these external resources into a Natural Language Processing (NLP) based geotagging model can aid in location expansion. CLAVIN-NERD already includes a gazetteer but expansion of the dataset of the gazetteer or incorporating a new gazetteer with a wider geographical coverage with knowledge graph information would enrich location support expansion and enable CLAVIN-NERD to handle diverse locations more effectively.

2. ***Contextual Word Embeddings:*** Context word embeddings (such as Word2Vec and BERT) have the ability to capture content and semantics effectively. In an NLP system, they can be useful as they allow the model to discern the context of a mention of locations which provides aid in the identification of location types - e.g. cities, businesses, and landmarks, beyond geotagging. A suggestion is to incorporate contextual word embeddings on location datasets to finely capture the location type and context.

3. ***Custom Entities:*** Named Entity Recognition (NER) models are trained to recognize general entities such as people, locations, organizations etc. In order to

achieve location support expansion, NER can be extended to be able to recognize precise entities which represent a specific location - e.g. public parks, bus stops, lakes etc. Incorporating this into CLAVIN-NERD would enable the model to identify a diversified set of locations in a more specialized way.

4. ***Ensemble Models:*** Ensemble models refer to a technique where multiple individual models are combined to improve the overall performance and robustness of the system. Using this method to combine multiple geotagging techniques and external data sources would enable the model to effectively handle different types of location references. Developing an ensemble model with integrated geotagging methods would improve location expansion accuracy and coverage.

5. ***Cross-lingual Transfer Learning:*** Cross-lingual transfer learning refers to the practice of leveraging knowledge from one language to improve the performance of the NLP model in another language. Transfer of knowledge across different languages is enabled through this technique, eliminating the need for extensive language training. By leveraging knowledge from pre-trained models in one language to another, CLAVIN-NERD would be able to identify location types in various languages, ensuring better location support expansion.

In conclusion, Location Support Expansion is a crucial advancement for CLAVIN-NERD. By leveraging external resources and adopting various techniques such as those mentioned above, the model can extract more comprehensive information about location, leading to better performance and accuracy.