



Report on Entity Recognition System for Toutche

- **Approach to Entity Recognition**

The entity recognition system developed for Toutche employs spaCy, a powerful and efficient library for natural language processing (NLP). The approach involves leveraging both pre-trained models and custom configurations to ensure the system can accurately identify entities relevant to customer inquiries. The primary steps in the development process included defining a list of entities specific to Toutche's domain, training the model using example phrases, and implementing an EntityRuler to enhance recognition accuracy for custom entities.

The use of spaCy's EntityRuler allows for the addition of patterns and keywords that are characteristic of the entities we wish to extract. This provides a robust way to incorporate Toutche-specific terminology, ensuring the model recognizes and extracts these entities effectively alongside general ones.

- **Handling Toutche-Specific Entities**

Toutche operates in a niche market involving electric bikes and customer support, which necessitates the recognition of specialized terminology. In defining the entities, we categorized them into four key groups: Product Names, Technical Issues, Customer Support Actions, and Locations. Each category was populated with example phrases reflecting real customer interactions, which were crucial for training the model.

For instance, entities like "Heileo H200" and "battery issue" were identified as critical for understanding customer inquiries. By integrating these specific terms into the training

dataset and the EntityRuler, the system is able to better understand context and meaning, resulting in improved accuracy for relevant queries.

- **System Performance on Test Queries**

The system was evaluated using a set of 20 test queries that represented typical customer support scenarios. Upon testing, the entity recognition system performed well, with a high accuracy rate in identifying and extracting the relevant entities. For instance, queries like "I recently purchased a Toutche bike" successfully extracted the entity "Toutche Bike," demonstrating the model's capability to recognize product names effectively.

However, there were some instances where technical issues arose, particularly with more complex phrases or when multiple entities were present in a single query. For example, queries containing both product names and technical issues occasionally led to partial recognition or misclassification. Overall, the system exhibited satisfactory performance, with room for refinement in handling overlapping entities.

- **Potential Improvements or Extensions**

To enhance the entity recognition system further, several improvements can be considered. Firstly, expanding the dataset with additional example phrases could help improve accuracy, particularly for less common queries. Training a custom spaCy model with domain-specific data may also yield better results for entity extraction.

Another potential extension involves implementing a feedback loop where customer support agents can review and provide feedback on entity recognition results. This feedback could then be used to refine and retrain the model, continuously improving its performance.

Additionally, integrating the entity recognition system with an intent classification model could create a more robust customer support solution, enabling more effective routing of inquiries and automating responses based on recognized entities.