NER-Entity-Mapper
Project Kickoff

Using NLP for Business News Entity Recognition & Linking
Presented by: Pablo Faundez-Garces
Date: February 8th, 2025

Project Overview

NER-Entity-Mapper is a data science project that extracts and links organizations mentioned in business news articles to structured databases (e.g., Wikipedia company lists). It uses Named Entity Recognition (NER) and Named Entity Linking (NEL) to resolve entity variations and improve accuracy.

Project Objectives

- Extract organizations from business news articles using NER

- Link entity mentions to structured databases (e.g., Wikipedia, stock market lists)

- Handle entity variations with fuzzy matching and knowledge graph techniques

- Provide structured data for business intelligence applications

- Develop an API for seamless integration

Project Phases

1 **PLANNING** - DEFINE PROJECT SCOPE, WORKFLOW, AND GUIDELINES

2 **DATA COLLECTION** - GATHER BUSINESS NEWS DATASETS AND STRUCTURED ENTITY DATABASES

3 **DATA PREPROCESSING** - CLEAN AND NORMALIZE DATASETS FOR MODEL TRAINING

4 **MODEL DEVELOPMENT** - BUILD AND TRAIN NER AND NEL MODELS

5 **EVALUATION & TESTING** - ASSESS MODEL PERFORMANCE AND REFINE APPROACHES

6 **DEPLOYMENT** - DEPLOY MODELS AND API FOR REAL-WORLD USE

7 **COLLABORATION & MAINTENANCE** - ESTABLISH CONTRIBUTION AND REVIEW PRACTICES

Key Pending Tasks

◆ IDENTIFY AND COLLECT RELEVANT BUSINESS NEWS DATASETS

◆ OBTAIN STRUCTURED ENTITY DATABASES (E.G., WIKIPEDIA COMPANY LISTS)

◆ DEFINE AND ESTABLISH GIT WORKFLOW AND CODE REVIEW GUIDELINES