Product Prediction Application - Delivery Document

Introduction

Welcome to the Product Prediction Application, a tool designed to predict furniture products for a given list of URLs. Leveraging advanced natural language processing techniques, this application analyzes web page content to determine the presence of furniture products.

Key Features

1. Product Prediction

- **Objective**: Predict whether a given target web URL contains furniture products.

2. Data Collection

- **Objective**: Collect training data from a predefined shortlist of furniture stores provided in a CSV file (`short_furniture_stores.csv`).

3. Model Training

- **Objective**: Process collected data and train a BERT (Bidirectional Encoder Representations from Transformers) Named Entity Recognition (NER) model. Here, we create 'PRODUCT' entity to train model with collected data.

- **Result**: The trained model is capable of identifying product-related entities within the text.

Building Model

1. Prepare Training Data

(1) A shortlist of furniture stores is provided in a CSV file (`short_furniture_stores.csv`).

(2) The collect_and_save_data(url_short_list) method collects data from given urls and executes several cleaning and preprocessing task.

(3) Collected and cleaned data will be saved as training data in a file (`training_data.csv`).

2. Train the Model

(1). We define BERT NER dataset class in ner_dataset.py file.

(2). By using training data get_train_data() method creates entities list data as shown below:

```
  train_data = [
      ('bathroom mirrors', {'entities': [(9, 15, 'PRODUCT')]}),
      ('bookcases and library units',    {'entities': [(0, 7, 'PRODUCT')]}),
      ….
  ]
```

(3). Entities list data is used to train model and saved in the `ner_model` folder for future predictions.

(4). If the model is already built, the application skips this step.

#1 and #2 is performed for first to build the model. Once model is built the application skips those steps in execution.

Product Prediction

Predict Products

1. Provide a list of target web URLs in CSV format (`furniture stores pages.csv`).

2. The predict_products() method conducts prediction by using the trained model and and given url. It includes the following steps:

- Scraping web by using URL.
- Collecting the content of page.
- Cleaning and process the collected data
- Predicting the collected text
- Print the result if product is predicted.

Result

The application provides results in the following format:

```
PRODUCT available in this page! URL: https://www.factorybuys.com.au/products/euro-top-mattress-king
PRODUCT available in this page! URL: https://dunlin.com.au/products/beadlight-cirrus
PRODUCT available in this page! URL: https://themodern.net.au/products/hamar-plant-stand-ash
PRODUCT available in this page! URL: https://interiorsonline.com.au/products/interiors-online-gift-card
```

Usage

1. Clone provided repository to your local machine:

2. Navigate to the project directory:

```
cd your-repository
```

3. Install the required dependencies:

```
pip install -r requirements.txt
```

4. Follow the usage instructions to prepare training data, train the model, and predict products.

5. Run the application:

```
python main.py
```

Appendix

A. The content of ‘short furniture stores pages.csv’ file

https://www.factorybuys.com.au/products/euro-top-mattress-king
https://dunlin.com.au/products/beadlight-cirrus
https://themodern.net.au/products/hamar-plant-stand-ash
https://furniturefetish.com.au/products/oslo-office-chair-white
https://hemisphereliving.com.au/products/
https://home-buy.com.au/products/bridger-pendant-larger-lamp-metal-brass
https://interiorsonline.com.au/products/interiors-online-gift-card
https://beckurbanfurniture.com.au/products/page/2/
https://livingedge.com.au/products/tables/dining
https://edenliving.online/collections/summerloving/products/nice-lounge-1

B. The content of ‘short furniture stores pages.csv’ file

https://www.factorybuys.com.au/products/euro-top-mattress-king
https://dunlin.com.au/products/beadlight-cirrus
https://themodern.net.au/products/hamar-plant-stand-ash
https://furniturefetish.com.au/products/oslo-office-chair-white
https://hemisphereliving.com.au/products/
https://home-buy.com.au/products/bridger-pendant-larger-lamp-metal-brass
https://interiorsonline.com.au/products/interiors-online-gift-card
https://beckurbanfurniture.com.au/products/page/2/
https://livingedge.com.au/products/tables/dining
https://edenliving.online/collections/summerloving/products/nice-lounge-1

C. The content of ‘training_data.csv’ file

aaron mid century modern upholstered sofa
aaron retro sofa
about the furniture
accent chairs
accent occasional chairs
accent pillows
accent tables
accessories
adirondack accessories
adirondack chairs
adirondack chairs and sets
adirondack chairs under
adirondack cushions