GenBank CDS Parser and Database Manager

This Python script parses GenBank files, extracts coding sequences (CDS) features, and manages them in a tabular database. It provides functionalities to list, insert, remove, and search entries within the database. The database is saved as a CSV file for storage.

PREREQUISITES

This script requires the following packages:
- pandas
- biopython
- prettytable

You can install them using pip install

USAGE

Run the script. Follow the on-screen prompts to interact with the database.

"Enter GenBank file path: " - Provide the path to the GenBank file you want to parse. Ensure the GenBank file is correctly formatted and contains CDS features. The script parses the given GenBank file to extract CDS features and returns a pandas DataFrame containing the locus tag, product, and protein ID of each CDS.

"Enter filename to save database (e.g., database.csv): " - Provide a filename to save the initial database as a CSV file.

"Enter your choice (1-5): " - Provide the number of your option of choice from the following:

1. **"List Database"** - Displays all entries in the DataFrame in a PrettyTable format

2. **"Insert New Entry" -** Inserts a new entry into the DataFrame with the locus tag, product and protein ID provided by the user and saves the updated DataFrame to a CSV file.

3. **"Remove Entry" -** Removes an entry with the user-specified locus tag from the DataFrame and saves the updated DataFrame to a CSV file

4. **"Search Entry"** - Searches for entries containing the user-specified keyword in any column and displays the results in a PrettyTable format.

5. **"Exit"** - Closes the program.

NOTE

The CSV file will be overwritten each time an entry is added or removed.