Priya .K

Task 2

```
pip install requests beautifulsoup4
```

```
Requirement already satisfied: requests in c:\users\hp\anaconda3\lib\site-packages (2.32.2)
Requirement already satisfied: beautifulsoup4 in c:\users\hp\anaconda3\lib\site-packages (4.12.3)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\hp\anaconda3\lib\site-packages (from requests) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\hp\anaconda3\lib\site-packages (from requests) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\hp\anaconda3\lib\site-packages (from requests) (2.2.2)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\hp\anaconda3\lib\site-packages (from requests) (2024.8.30)
Requirement already satisfied: soupsieve>1.2 in c:\users\hp\anaconda3\lib\site-packages (from beautifulsoup4) (2.5)
Note: you may need to restart the kernel to use updated packages.
```

```python
def main():
    conn = init_db()

    # Fetch homepage
    homepage_html = fetch_url(BASE_URL)

    # Extract article links
    article_links = extract_article_links(homepage_html)

    for link in article_links:
        print(f"Processing: {link}")
        try:
            # Fetch article page
            article_html = fetch_url(link)

            # Extract details
            article_details = extract_article_details(article_html)

            # Save to database
            save_article_to_db(conn, article_details)

        except Exception as e:
            print(f"Failed to process {link}: {e}")

    conn.close()
    print("Scraping completed!")

if __name__ == "__main__":
    main()
```

```
Processing: https://indianexpress.com/article/business/economy/low-high-skilled-jobs-gap-rising-as-manufacturing-stagnation-continues-9569727/
Processing: https://indianexpress.com/article/business/aviation/india-has-immense-potential-for-regional-planes-looking-at-it-as-our-main-market-atr-cco-alexis-vidal-9569720/
Processing: https://indianexpress.com/section/business/
Processing: https://indianexpress.com/article/business/amid-global-transformer-supply-crunch-green-energy-developers-face-grid-access-delays-9571594/
Processing: https://indianexpress.com/article/business/aviation/air-india-refit-programme-legacy-aircraft-9572368/
Processing: https://indianexpress.com/article/business/pralhad-joshi-says-rs-32-45-lakh-crore-finance-pledged-for-renewables-sector-9570487/
Processing: https://indianexpress.com/article/business/shapoorji-pallonji-afcons-sebi-nod-rs-7000-crore-ipo-9571159/
Processing: https://indianexpress.com/article/business/bajaj-housing-lists-at-114-per-cent-premium-market-cap-9571051/
Processing: https://indianexpress.com/article/business/sebi-backtracks-withdraws-its-staff-misguided-by-external-elements-press-release-9570656/
Processing: https://indianexpress.com//section/business/
Processing: https://indianexpress.com/article/business/market/fpis-reverse-strategy-buy-more-equities-via-exchanges-on-us-rate-cut-anticipation-9569719/
Scraping completed!
```

```python
import sqlite3

# Function to initialize the SQLite database and create the schema
def create_database_schema(db_name):
    """
    Creates a SQLite database with the specified schema.

    Args:
    db_name (str): The name of the SQLite database file.
    """
    try:
        # Connect to the SQLite database (or create it if it doesn't exist)
        conn = sqlite3.connect(db_name)
        cursor = conn.cursor()

        # Create the articles table if it doesn't already exist
        cursor.execute('''
            CREATE TABLE IF NOT EXISTS articles (
                id INTEGER PRIMARY KEY AUTOINCREMENT,
                title TEXT NOT NULL,
                author TEXT,
                publication_date TEXT,
                content TEXT NOT NULL
            )
        ''')

        # Commit the changes and close the connection
        conn.commit()
        print(f"Database schema created successfully in '{db_name}'.")

    except sqlite3.Error as e:
        print(f"An error occurred while creating the database schema: {e}")

    finally:
        if conn:
            conn.close()

# Usage example
if __name__ == "__main__":
    create_database_schema("articles.db")
```

```
Database schema created successfully in 'articles.db'.
```