Task 1: Collect and Structure Conference Artifact Data

1. Identify Conferences with Artifact Tracks:

Search for up to 50 relevant conferences in software engineering and related fields that feature artifact evaluation tracks.

Access their paper submission pages where artifact evaluation details are outlined, such as the guidelines, evaluation criteria, and badges.

2. Download Submission Pages:

For each identified conference, download the submission guidelines and any associated documents (HTML pages or downloadable files).

Keep a clean and organized folder structure where each conference has its folder containing all related downloaded data.

3. Convert Data to a Structured Format:

For each conference, extract key details from the submission pages:

Badges: Types of badges awarded (e.g., "Reusable," "Available") and their descriptions.

Guidelines: The criteria and guidelines provided for artifact evaluation.

Evaluation Process: Any details on the stages or methods of evaluation (e.g., how artifacts are assessed).

Convert this information into a structured format (e.g., a JSON or CSV file for each conference) for consistency. Each structured file should include fields like "conference name," "year," "badge types," and "evaluation criteria."

4. Create Two Datasets:

Original Dataset: The downloaded HTML pages or documents of submission guidelines for each conference, stored in their respective folders.

Structured Dataset: A standardized, structured representation of the guidelines and badges for each conference, consolidated in a format suitable for analysis (e.g., JSON or CSV files).

Task 2: Gather Repository Data from Artifact Tracks

1. Find Repositories Linked to the Conferences:

For each conference that has an artifact track, gather the links to repositories associated with accepted artifacts. These links are often found on the conference's artifact track pages or related documentation.

2. Document Repository Links:

Create a Link-only Dataset: Organize the repository links in a structured manner without cloning the repositories. Each conference will have its folder, and within each folder, there will be a file listing all the repository URLs.

Add Metadata: Along with the repository link, include any additional metadata such as the repository name, associated conference, and year.

3. Create Two Folders for Each Conference’s Repositories:

Links Folder: Contains a file listing all repository URLs for the conference (e.g., a text file or CSV with fields "repository name," "URL," "conference name," "year").

Metadata Folder: Optionally, include a separate file for more detailed metadata about each repository if available (e.g., artifact type, authors).