

SK네트웍스 Family AI 캠프 1기

-데이터 전처리-

1. 수집된 데이터 정의

a) 사용자의 깃허브 레포지토리

- 사용자의 깃허브에서 레포지토리 목록을 가져옵니다.

b) 사용자의 깃허브 레포지토리 브랜치

- 사용자의 깃허브의 레포지토리 브랜치 목록을 가져옵니다.

c) 사용자의 깃허브 레포지토리 브랜치 커밋

- 사용자의 깃허브 레포지토리 브랜치에 있는 커밋 목록을 가져옵니다.

d) 생성된 백로그

- 사용자의 깃허브 레포지토리를 기반으로 LLM이 생성한 백로그를 저장합니다.

e) 사용자가 저장한 결과 보고서

- 백로그를 기반으로 생성한 결과 보고서를 저장합니다.

2. 데이터 전처리

- 커밋 목록 저장 시, 필요한 내용만 가져오도록 전처리
Example)

```
[
  {
    "url": "https://api.github.com/repos/octocat/Hello-World/commits/6dcb09b5b57875f334f61aebd695e2e4193db5e",
    "sha": "6dcb09b5b57875f334f61aebd695e2e4193db5e",
    "node_id": "MDY6Q29tbWl0NmRjYjA5YjViNTc4NzVmMzH0ZjYxYVViZWQ2OTVlMmU0MTkzZGI1ZQ==",
    "html_url": "https://github.com/octocat/Hello-World/commit/6dcb09b5b57875f334f61aebd695e2e4193db5e",
    "comments_url": "https://api.github.com/repos/octocat/Hello-World/commits/6dcb09b5b57875f334f61aebd695e2e4193db5e/comments",
    "commit": {
      "url": "https://api.github.com/repos/octocat/Hello-World/git/commits/6dcb09b5b57875f334f61aebd695e2e4193db5e",
      "author": {
        "name": "Monalisa Octocat",
        "email": "support@github.com",
        "date": "2011-04-14T16:08:49Z"
      },
      "committer": {
        "name": "Monalisa Octocat",
        "email": "support@github.com",
        "date": "2011-04-14T16:08:49Z"
      },
      "message": "Fix all the bugs",
      "tree": {
        "url": "https://api.github.com/repos/octocat/Hello-World/tree/6dcb09b5b57875f334f61aebd695e2e4193db5e",
        "sha": "6dcb09b5b57875f334f61aebd695e2e4193db5e"
      },
      "comment_count": 0,
      "verification": {
        "verified": false,
        "reason": "unsigned",
        "signature": null,
        "payload": null
      }
    },
    "author": {
      "login": "octocat",
      "id": 1,
      "node_id": "MDQ6VXNlcjE=",
      "avatar_url": "https://github.com/images/error/octocat_happy.gif",
      "gravatar_id": "",
      "url": "https://api.github.com/users/octocat",
      "html_url": "https://github.com/octocat",
      "followers_url": "https://api.github.com/users/octocat/followers",
      "following_url": "https://api.github.com/users/octocat/following{/other_user}",
      "gists_url": "https://api.github.com/users/octocat/gists{/gist_id}",
      "starred_url": "https://api.github.com/users/octocat/starred{/owner}/{repo}",
      "subscriptions_url": "https://api.github.com/users/octocat/subscriptions",
      "organizations_url": "https://api.github.com/users/octocat/orgs",
      "repos_url": "https://api.github.com/users/octocat/repos",
      "events_url": "https://api.github.com/users/octocat/events{/privacy}",
      "received_events_url": "https://api.github.com/users/octocat/received_events",
      "type": "User",
      "site_admin": false
    },
    "committer": {
      "login": "octocat",
      "id": 1,
      "node_id": "MDQ6VXNlcjE=",
      "avatar_url": "https://github.com/images/error/octocat_happy.gif",
      "gravatar_id": "",
      "url": "https://api.github.com/users/octocat",
      "html_url": "https://github.com/octocat",
      "followers_url": "https://api.github.com/users/octocat/followers",
      "following_url": "https://api.github.com/users/octocat/following{/other_user}",
      "gists_url": "https://api.github.com/users/octocat/gists{/gist_id}",
      "starred_url": "https://api.github.com/users/octocat/starred{/owner}/{repo}",
      "subscriptions_url": "https://api.github.com/users/octocat/subscriptions",
      "organizations_url": "https://api.github.com/users/octocat/orgs",
      "repos_url": "https://api.github.com/users/octocat/repos",
      "events_url": "https://api.github.com/users/octocat/events{/privacy}",
      "received_events_url": "https://api.github.com/users/octocat/received_events",
      "type": "User",
      "site_admin": false
    },
    "parents": [
      {
        "url": "https://api.github.com/repos/octocat/Hello-World/commits/6dcb09b5b57875f334f61aebd695e2e4193db5e",
        "sha": "6dcb09b5b57875f334f61aebd695e2e4193db5e"
      }
    ]
  }
]
```

위와 같은 출력에서, 커밋 내용만 가져오도록 전처리 (코드 링크)