

2024.07.02~03

뉴스크롤링후 분산처리해서
wordcount api해서 특정 wordcloud

크롤링해보기

크롤러짜보기

덕디비--캐시...

RDBMS 파싱해서 하둡에 넣고 하둡에서 다시 WORDCOUNT해서 WORDCLOUD해보기...

네이버 기사 크롤링

메인섹션
서브섹션

db- > hadoop 에 넣자
하루단위로 읽어서 하루씩 넣자
csv말고 parquet

<https://ohmyz.sh/#install>

이것도 깔았고..

sudo apt git flow도 깔았고

git flow init

git flow feature start mort 했고

git status

git branch로 확인가능하다...

git add 하고 commit 해서 이메일 아이디정하

git push --set-upstream origin feature/Mort

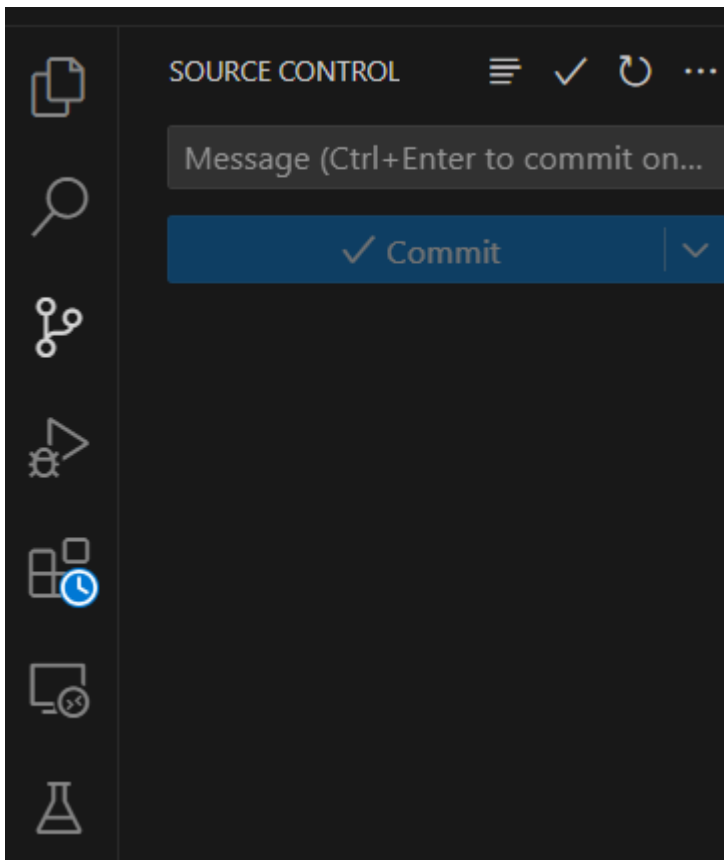
내 이메일 깃허브 비번치는데

ide가 아니므로 vscode가서

.gitignore 추적금지...

.env로 중요한보안파일을 옮기고 import해서 사용한다...

깃허브 setting에 repository 에 secert scanning alerts "security"에 있따..



3번째 눌러서... source control 해서 commit한 다음에 그리고 주소를 통하여 github 들어가면
일단 vscode와 내 git을 연동하고, accept하 review assign 다 설정한 다음 확인을 누르면 누군가
가가 approve해준다..

스파크로 읽어와서 wordcount
mysql에 table만들어서 넣고 erd생각

컨트롤마우스누르면 파일주소알수있