Kravspecifikation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Navn | Download og Statusregistrering af PDF-rapporter fra GRI-rapportering (2017-2020) | | | |
| Opgave | Udvikling af et Python-program, der downloader PDF-rapporter fra et givent datasæt, gemmer status for hver rapport, og opdaterer metadata om downloadet status og filstørrelse. Programmet skal kunne håndtere ustabile internetforbindelser og fejlhåndtere alternative download-links. | | | |
| Krav/prioriteter: *Her kan du oplyse de vigtigste elementer i dit software* | Håndtering af primært og alternativt link: Hvis download fra det primære link fejler, prøv det alternative link.  Stabilitet og ydeevne: Programmet skal være stabilt og hurtigt, og håndtere op til 10 downloads samtidigt via multithreading.  Download-status: Status skal logges som "Downloadet" eller "Ikke downloadet" og opdateres i en status-fil (download\_status.xlsx).  Navngivning af filer: Navngiv hver downloadet PDF efter værdien i kolonnen "BRnum."  Filstørrelse: Registrer og opdater hver rapports filstørrelse i download-status-filen.  Aflevering: Det færdige program og listen over downloadede filer leveres til kunden, som skal kunne gentage processen fremover. | | | |
| Kodesprog | Python | | | |
| Evt. frameworks | Pandas: Til datahåndtering  Requests: Til at hente PDF-filer via HTTP  ThreadPoolExecutor (concurrent.futures): Til multithreading  TQDM: Til visuel fremdriftsindikator i terminalen  Tabulate: Til tabelvisning af resultater i terminalen  Socket: Til kontrol af internetforbindelse | | | |
| Estimeret tid | 10-14 timer | | | |
|  | | | | |
| Featurelist *Her kan du tilføje hvilket features som skal kodes til dit software* | Automatisk internetforbindelsestjek: Tjekker aktiv internetforbindelse før og under download, med genoptagelse efter midlertidige udfald.  Multithreaded download: Håndterer op til 10 downloads samtidig, hvilket forbedrer ydeevnen.  Alternativt link-faldback: Prøver et alternativt link, hvis primært link ikke virker.  Download-status og filstørrelse: Opdaterer status og filstørrelse i en Excel-fil for hver rapport.  Brugervenlig grænseflade: Viser downloadstatus og fremdrift i terminalen med brug af progress bars og tabeller.  Automatisk oprettelse af download-mappe: Sikrer, at mappen til PDF-filer eksisterer, eller opretter den automatisk. | | | |
| Work breakdown *Her kan du bryde dit software ned i arbejdsopgaver og estimere tiden det vil tage at bygge hvert et element* | Arbejdsopgave | Start dato | Forventet slutdato | Tidsforbrug |
| Initial analyse og planlægning | 04-11-2024 | 04-11-2024 | 1 time |
| Opsætning af miljø og biblioteker | 04-11-2024 | 04-11-2024 | 0.5 timer |
| Læsning og fix af Excel-datasæt | 04-11-2024 | 04-11-2024 | 1 time |
| Internetforbindelsestjek og håndtering | 04-11-2024 | 08-11-2024 | 0.5 timer |
| Primært og alternativt download-link | 04-11-2024 | 08-11-2024 | 2 timer |
| Multithreading implementering | 04-11-2024 | 04-11-2024 | 1 time |
| Statusopdatering og logning | 04-11-2024 | 04-11-2024 | 0.5 timer |
| Tabelvisning og progress bar | 04-11-2024 | 04-11-2024 | 0.5 timer |
| Test og fejlretning | 04-11-2024 | 08-11-2024 | 2 timer |
| Dokumentation og aflevering | 08-11-2024 | 08-11-2024 | 4 timer |
|  |  |  |  |
|  |  |  |  |