Data Engineer Technical Test

1. Buatlah sebuah forward proxy dalam bahasa pemrograman pilihanmu. Forward proxy ini harus:
    a. Bind ke port 9919 di localhost
    b. Dapat digunakan untuk mengeksekusi command berikut:

No	Command	Output	Remarks
1	curl -x http://localhost:9919 https://google.com/search -vvv	1. HTTP 2xx	
2	curl -x http://localhost:9919 https://en.wikipedia.org/wiki/Proxy_server -vvv	1. HTTP 2xx	
3	curl -x http://localhost:9919 https://en.wikipedia.org/wiki/Proxy_server -vvv	1. HTTP 2xx 2. Hapus setiap kemunculan kata **software**, dan di akhir dari http response body sertakan jumlah kata **software** yang dihapus	Bonus

2. Buat sebuah wikipedia scraper dalam bahasa pilihanmu. Scraper ini harus bisa dijalankan via bash script. Dimisalkan bash scriptnya adalah run_scraper.sh, maka command yang harus bisa memenuhi fungsi berikut:
    a. ./run_scraper.sh [phrase]. Melakukan pencarian dan meng-ekstrak hasil pencarian menjadi sebuah json file yang formatnya sesuai lampiran di halaman berikut ini. Json file disimpan di direktori yang sama dengan bash script dieksekusi. Jika command ini dilakukan berkali-kali, json file tidak boleh menimpa file yang sudah ada sebelumnya. Pencarian harus memungkinkan lebih dari 1 kata
    b. ./run_scraper [phrase] [proxy_url]. Sama seperti poin (a), akan tetapi ada jika dimasukkan argumen [proxy_url], maka eksekusi scraping harus menggunakan proxy_url sebagai forward proxy

3. (Bonus) Buatlah sebuah wikipedia scraper yang memungkinkan:
    a. Scraper diberikan sekumpulan link awal artikel wikipedia untuk discraping, misalnya https://en.wikipedia.org/wiki/Proxy_server. Kemudian, link

tersebut di-ekstrak judul, body artikel, link-link artikel terkait, serta tanggal pembuatan artikel

b. Kemudian, untuk setiap link-link artikel terkait, scraper harus bisa melakukan fungsi pada poin (a), namun scraper harus bisa mendeteksi link mana yang sudah di-scrap, mana yang belum. Scraper tidak boleh melakukan scraping link yang sama 2x
c. Proses scraping berhenti jika:
    i. Proses di-kill, atau dihentikan oleh pengguna
    ii. Tidak ada lagi link yang bisa di-scrap

Sertakan batasan dan asumsi yang kamu gunakan dalam membangun wikipedia scraper ini.

Ketentuan pengerjaan:

1. Boleh menggunakan library milik orang lain, namun khusus untuk menangani logic yang terkait scraping dan forward proxy tidak boleh menggunakan library buatan orang lain

Deliverable:

1. Github link sebagai repository hasil pengerjaan
2. (Bonus) screencast eksekusi no (1), (2), dan (3)

Struktur json:

```
[{
    "title":"Some title",
    "link": "https://some-link",
    "content": "Some long content",
    "createdAt": "yyyy-MM-DDTHH:mm:ss.ZZZ",
    "category": "category of the article, if any"
}]
```