Gelombang 2 - Regex [37 poin]

HTML parsing: https://s.id/u56EN. Gunakan fitur *inspect element* untuk melihat struktur halaman. Buka webpage menggunakan private/incognito mode jika terhalang *paywall*.

---

1. **[11 poin] Tulis regex** untuk mengambil semua paragraf, i.e. regex anda harus mampu mengambil:

   [**1:** “As we all know, smartphones are catching up to the lifestyles of its users. That’s no secret, users seeking every solution in their phones whether to book a cab or to find the nearest good restaurant. To get things done easily, people have many reasons to expect their mobile experiences will be good as PC based ones, but most of the webs have failed.”, **2:** “According to Google, 53% of users bounce off a web page which takes more than eight seconds. And, yet the average homepage still takes 15 seconds to load on mobile phones.”, ...]

   Namun harus mampu menghindari paragraf judul, seperti:

   > **Before we mover further, get an idea about the term Progressive Web apps and responsive web apps-**
   > **Progressive Web Apps -**
   > Etc.

   Jelaskan bagaimana regex anda membedakan 2 kasus di atas.

   Hint: Gunakan grup. Tuliskan regexnya saja, tidak perlu tulis kode untuk memproses grup tersebut lebih lanjut. Anda boleh menuliskan dua regex, 1 untuk mengambil semua paragraf (termasuk paragraf judul), dan 1 lagi untuk mengambil khusus paragraf judul. Anda juga bisa menuliskan solusi yang memerlukan satu regex saja. Untuk soal ini, tidak masalah jika paragraf anda mengandung HTML tag lagi di dalamnya (seperti link).

---

2. **[11 poin]** Dari soal 1.b., anda sekarang sudah memiliki teks yang kemungkinan memiliki link tag, seperti:

   This was one of the reasons that the idea of progressive web apps was proposed. Google has introduced a new term in 2015 referred to as progressive web apps with a thought to provide native-like experiences to the users visiting a website.

   yang HTML code-nya adalah:

```
This was one of the reasons that the idea of <a
href="https://blogs.emorphis.com/progressive-web-apps/" class="et ib"
rel="noopener nofollow">progressive web apps</a> was proposed. Google has
introduced a new term in 2015 referred to as progressive web apps with a
thought to provide native-like experiences to the users visiting a
website.
```

Tulis regex untuk menghapus tag link tersebut sehingga teksnya mejadi bersih: "*This was one of the reasons that the idea of progressive web apps was proposed. Google has introduced a new term in 2015 referred to as progressive web apps with a thought to provide native-like experiences to the users visiting a website.*"

Hint: gunakan 3 grup. Tuliskan regexnya saja, tidak perlu tulis kode untuk memproses 3 grup tersebut lebih lanjut. **Jelaskan bagaimana regex anda bekerja** (jika anda mengikuti cara ini, jelaskan peran dari masing-masing dari 3 grup tersebut).

---

3. **[Total 15 poin] Singkatan dan kepanjangannya.**
    a. **[5 poin]** Tulis regex yang *match* dengan semua singkatan yang muncul di HTML page tersebut. Singkatan yang dimaksudkan di sini adalah yang bertipe inisialisme (contoh: CPU, CPUs, dan C.P.U.), namun tidak termasuk yang bertipe abreviasi atau akronim (Mr., Dr., Prof., etc. tidak perlu dipedulikan). Silakan berasumsi bahwa semua huruf kapital berturut-turut merupakan bentuk inisialisme (misal, "HAHA", kita anggap juga singkatan).
    b. **[10 poin]** Kemudian, tulis *pseudocode* untuk suatu function/method yang menerima input suatu inisialisme dan mengeluarkan output suatu regex untuk mencari semua kemungkinan kepanjangan dari singkatan tersebut dari teks. Misal, untuk input "PWA", maka regexnya harus match dengan "progressive web apps", "permanent worker association", etc..

Hint: anda dapat menggunakan fungsi bantu untuk mengakses dan memanipulasi string, seperti charAt, startsWith, endsWith, lowercase, uppercase, append, etc