Headers uit HTML 2 (haakjes)

In de vorige opdracht heb je geleerd hoe je headers uit een HTML tekst kan halen.

Verander de code nu zodat je alleen de header afdrukt.

In de vorige opdracht kreeg je als resultaat bijvoorbeeld:

> *<h1>Dit is een header</h1>*

Pas de code nu aan zodat je alleen

> *Dit is een header*

afdrukt.

Deze keer moet je zelf op zoek gaan hoe dat moet. Gebruik het internet en eventueel ChatGPT. Er zijn meerdere oplossingen mogelijk, maar je moet de door jouw gekozen oplossing wel begrijpen en kunnen uitleggen.

Tip

Met haakjes kan je aangeven welk deel van de reguliere expressie (regex) je in het resiltaat wilt zien.

Als je in het pattern van

```
re.findall(pattern, html_text)
```

haakjes gebruikt om aan te geven welk deel je als resultaat wilt, dan krijg je in het resultaat alleen dat deel dat tussen haakjes staat.

voorbeeld

```
text = "Vandaag is het 20-07-2023. Mijn verjaardag is op 01-05-1985 en mijn jubileum is op 12-10-2010."
pattern = r"\d{2}-\d{2}-(\d{4})"
matches = re.findall(pattern, text)
```

In dit voorbeeld wordt gezocht naar een patroon dd-dd-dddd en alleen de laatste dddd wordt als match gevonden.

Er worden in dit voorbeeld dus drie matches gevonden; 2023, 1985 en 2010.

Opdracht

Pas de code van de vorige opdracht aan zodat je de <h1> en </h1> niet meer afdrukt als onderdeel van het resultaat. Zie hierboven een voorbeeld.

Gebruik internet, ChatGPT ,.... maar zorg dat je de oplossing begrijpt en kan uitleggen.

Inleveren

1. Aangepaste code en zet in commentaar in je code een uitleg in je eigen woorden van hoe het werkt.
2. Laat zien in een screen dump dat je code werkt.

Revision #3
Created 2023-07-31 13:12:03 UTC by Max
Updated 2023-07-31 13:52:14 UTC by Max