

Bronco ID: | 0 | 1 | 5 | 2 | 6 | 2 | 6 | 2 | 4 |

Last Name: Nguyen

First Name: Loc

1a. 3 – 5 – 9 – 2 – 11 – 4 – 10 – 8 – 7 – 12

1b. Pages 1 and 6 would be located on the deep web because the seed URL of page 3 has no link pathway to access those two pages, which wouldn't be indexed.

1c. 3 – 5 – 9 – 1 – 2 – 11 – 6 – 4 – 10 – 8 – 7 – 12

1d. None of the pages will be located on the deep web with a new pathway added from seed URL of 3 to 1. This is because all the pages are reachable and will be indexed, leaving none that aren't indexed so nothing will be in the deep web.

```
2a. print(bs.title.get_text())
2b. print(bs.ol.get_attribute_list(", 'To my friends'))
2c. print(bs.find('tr', {'class': 'tutorial2'}).get_text())
2d. for text in bs.find_all('h2', string=re.compile("tutorial")):
    print(text.get_text())
2e. for text in bs.find_all("", string=re.compile("HTML")):
    print(text.get_text())
2f. print(bs.find('tr', {'class': 'tutorial1'}).get_text())
2g. for image in bs.find_all('img'):
    print(image['src'])
```

- 3a. String “c” and “bac” will match  $b?a?c$  (start with 0 or 1 b, then 0 or 1 a, and end with c)
- 3b. String “ccca” and “a” will match  $c^*a$  (start with 0 or more c, end with a)
- 3c. String “aca” and “acccca” will match  $ac^+a$  (start with a, 1 or more c, end with a)
- 3d. String “Aaa” and “Caaa” will match  $[A-Z]a\{2,3\}$  (start with A-Z, end with 2-3 a)
- 3e. String “cac” and “cbc” will match  $c(a|b)c$  (start with c, pick a or b, end with c)
- 3f. String “0123456” and “98989898” will match  $\backslash d\{2,2\}\backslash \backslash d\{2,2\}\backslash \backslash d\{2,4\}$  (start with 2 numbers 0-9, then 2 numbers 0-9, end with 2-4 numbers 0-9)
- 3g. String “0ab9” and “10” will match  $\backslash d(ab)^*\backslash d$  (start with any number 0-9, 0 or more ab, and end with any number 0-9)
- 3h. String “cbba” and “cba” will match  $^cb^+a$  (start with c, 1 or more b, end in a)

4. <https://github.com/Skyhorizon2021/CS4250/blob/master/Assignment3/crawler.py>

5. <https://github.com/Skyhorizon2021/CS4250/blob/master/Assignment3/parser.py>