Bronco ID: 014644097

Last Name: Pham

First Name: Nathan

* 1. 3-5-9-2-11-4-10-8-12-7
  2. Webpages that are located on the deep web are pages that require credential to access or do not have any links pointing to that webpage. In this example using page 3 as the seed URL, page 1 is on the deep web because there are no links pointing to it.
  3. 3-5-9-1-2-11-6-4-10-8-12-7
  4. Because all the pages have a link pointing to it, there are no webpages that are on the deep web.
  5. bs.title
  6. bs.find\_all('li')[3])
  7. bs.find\_all('tr')[1].find\_all('td')
  8. bs.find\_all('h2', text = re.compile('.\*tutorial.\*'))
  9. text = re.compile('.\*HTML.\*')
  10. bs.find('tr').find\_all('td'))
  11. bs.table.find\_all('img')
  12. bc, c
  13. ccccca, a
  14. aca, accca
  15. Ba3, Fa2
  16. cba, cbc
  17. 32/72/53, 92/42/34
  18. 12, 6ababababab9
  19. cbbbbbbbba, cbbba

<https://github.com/nathanvpham/CS4250Assignment3>

<https://github.com/nathanvpham/CS4250Assignment3>

file: parsing.py