



# Semester Project

CHAT BOT PROGRAM

CSCE 240

ANTON AARNINK





# Scope



## Representative William M. "Bill" Chumley

Republican - Spartanburg  
District 35 - Greenville & Spartanburg Counties - [Map](#)

### Columbia Address

326A Blatt Bldg.  
Columbia 29201

### Home Address

P.O. Box 22  
Reidville 29375

**Business Phone** (803) 212-6894

**Home Phone** (864) 303-2726

**Business Phone** (864) 303-2726

[Send message to Representative Chumley](#)

## Personal Information

- Landscaping & Farmer
- Residing at 3303 Greenpond Rd., Woodruff
- Born September 24, 1947 in Spartanburg
- Son of John W., Jr. and Dorothy D. Chumley
- Attended Spartanburg Technical College, Clemson University
- September 2, 1967 married Faye Wood, 2 children, Rob and Case
- Teacher, vocational school
- PTO President
- Coach, little league
- Co-Founder, Reidville Historical Society
- Member, Sons of Confederate Veterans
- Holder of six U.S. patents and two trademarks
- South Carolina Nurseryman
- Deacon, Poplar Springs Baptist Church

## Committee Assignments

- [Agriculture, Natural Resources & Environmental Affairs](#)

## Sponsored Bills in the House

- Primary Sponsor: ☒ Yes ☐ No
- Search Session:

## Voting Record

- Search Session:

## Service In Public Office

- House of Representatives, 2011 - Present

# District 35



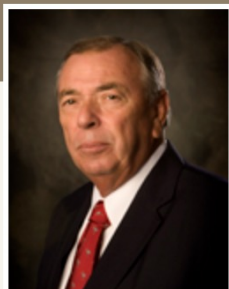
# Programing Language Choice

- ▶ Python
- ▶ Why?
  - ▶ Simple package integration
  - ▶ Wide number of packages available
  - ▶ Fast development
- ▶ What a different language could fix ...
  - ▶ Faster processing and run times





# Data



## Representative William M. "Bill" Chumley

Republican - Spartanburg  
District 35 - Greenville & Spartanburg Counties - [Map](#)

### Columbia Address

326A Blatt Bldg.  
Columbia 29201

**Business Phone** (803) 212-6894

### Home Address

P.O. Box 22  
Reidville 29375

**Home Phone** (864) 303-2726

**Business Phone** (864) 303-2726

[Send message to Representative Chumley](#)

## Personal Information

- Landscaping & Farmer
- Residing at 3303 Greenpond Rd., Woodruff
- Born September 24, 1947 in Spartanburg
- Son of John W., Jr. and Dorothy D. Chumley
- Attended Spartanburg Technical College, Clemson University
- September 2, 1967 married Faye Wood, 2 children, Rob and Case
- Teacher, vocational school
- PTO President
- Coach, little league
- Co-Founder, Reidville Historical Society
- Member, Sons of Confederate Veterans
- Holder of six U.S. patents and two trademarks
- South Carolina Nurseryman
- Deacon, Poplar Springs Baptist Church

## Committee Assignments

- [Agriculture, Natural Resources & Environmental Affairs](#)

## Sponsored Bills in the House

- Primary Sponsor: ☒ Yes ☐ No
- Search Session:

## Voting Record

- Search Session:

## Service In Public Office

- House of Representatives, 2011 - Present

# What data is available?



Representative William M. "Bill" Chumley  
-  
Republican - Spartanburg-  
District 35 -  
Greenville & Spartanburg Counties  
-  
Map-  
326A Blatt Bldg.  
Columbia 29201-  
Business Phone  
(803) 212-6894-  
P.O. Box 22  
Reidville 29375-  
Home Phone  
(864) 303-2726-  
Business Phone  
(864) 303-2726-  
-  
li: Landscaping & Farmer  
li: Residing at 3303 Greenpond Rd., Woodruff  
li: Born September 24, 1947 in Spartanburg  
li: Son of John W., Jr. and Dorothy D. Chumley  
li: Attended Spartanburg Technical College, Clemson University  
li: September 2, 1967 married Faye Wood, 2 children, Rob and Case  
li: Teacher, vocational school  
li: PTO President  
li: Coach, little league  
li: Co-Founder, Reidville Historical Society  
li: Member, Sons of Confederate Veterans  
li: Holder of six U.S. patents and two trademarks  
li: South Carolina Nurseryman  
li: Deacon, Poplar Springs Baptist Church  
li: Agriculture, Natural Resources & Environmental Affairs  
li: House of Representatives, 2011 - Present

# What is retrieved from program ?



# Code Organization



# PA 1

- ▶ Packages used:
  - ▶ lxml – parsing engine
  - ▶ requests – web requests
  - ▶ BeautifulSoup4 – HTML parser
- ▶ This was a basic web scraping program

```
import requests
from bs4 import BeautifulSoup
from loadingAnimation import loading
import time
import os

def WebScraper():
    start = time.time()
    loading()
    path = path = os.getcwd().removesuffix("/src")
    with open(path + "/data/output.txt", "w") as file:
        file2 = open(path + "/data/stats.txt", "a")
        file2.truncate(0)

        url = "https://www.scstatehouse.gov/member.php?code=351704504"

        page = requests.get(url)

        soup = BeautifulSoup(page.content, 'lxml')

        name = soup.find("h2", class_="barheader").get_text().strip()

        file.write(name + '\n')

        # removes unused html so results later do not include
        soup.find('div', id='menu').decompose()
        soup.find('div', id="search").decompose()
        soup.find('div', id="submenu").decompose()
        soup.find('div', id="sidebar").decompose()

        soup.find('div', id="footer").decompose()
        soup.find('div', id="printfooter").decompose()
        for dat in soup(['form', 'script']):
            # Remove tags
            dat.decompose()

        for data in soup.find_all("p"):
            file.write(data.get_text(separator='\n').strip() + '- \n')

        for data2 in soup.find_all("li"):
            line = "{0}: {1}".format(data2.name, data2.text)
            if line != "*Search Session*":
                file.write("{0}: {1}".format(data2.name, data2.text) + '\n')

        file.close()
        page.close()
        end = time.time()
        runtime = round(end - start, 2)
        str1 = "***It took"
        str2 = "seconds to get information from the internet.***"
        str3 = " ".join([str1, str(runtime), str2])
        file2.write(str3)
        file2.write("\n")
        file2.close()
```

# PA 2

- ▶ This program processed the output of the web scraping
- ▶ At this point in development the program asked the user to choose the desired information from a console menu

```
with open(path + "/data/output.txt", "r+") as input_file:
    with open(path + "/data/processed.txt", "w") as output_file:

        print("What district do you have questions about?")
        district = input()

        if district != "35":
            print("I do not have information on that district ?")
            exit()

        print("You have selected district: " + district)

        print(
            "What would you like to know? \n1.) Personal Information\n2.) Full Name\n3.) Service Record\n4.) Committee "
            "Assignments\n5.) Contact Information\n*****Q- to quit*****")
        choice = input()
```

## User Seen Menu

```
print(
    "What would you like to know? \n1.) Personal Information\n2.) Full Name\n3.) Service Record\n4.) Committee "
    "Assignments\n5.) Contact Information\n*****Q- to quit*****")
choice = input()
```



1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 2679, 26

- ```
personal_info_list = ["info", "about", "from"]
phone_info_list = ["phone", "number", "contact"]
name_list = ["name", "full name", "last name"]
committee_list = ["committee"]
service_list = ["service", "record", "long", "served"]
address_list = ["live", "where", "address"]
quit_cases = ["q", "quit", "exit"]
all_list = ["everything", "all"]
```

```
# Created by antonaarnink at 4/4/22

import sys
import time

def loading():
    print("Loading:")

    animation2 = ["10%", "20%", "30%", "40%", "50%", "60%", "70%", "80%", "90%", "100%"]
    animation = ["[#####]", "[#####]", "[#####]", "[#####]", "[#####]", "[#####]", "[#####]", "[#####]", "[#####]", "[#####]"]

    for i in range(len(animation)):
        time.sleep(0.2)
        sys.stdout.write("\r" + animation[i % len(animation)] + animation2[i % len(animation2)])
        sys.stdout.flush()

    print("\n")
```

# PA 5

- ▶ Improved key word search
- ▶ Consolidated and streamline code for processing and scraping
- ▶ Migrated some functions to the main program file

```
for element in personal_info_list:
    if re.search(element, ans, re.IGNORECASE):
        print(personal_info)
        for el in personal_info:
            file.write(el)
            file.write("\n\n")
        sys.stdout.flush()
for element in phone_info_list:
    if re.search(element, ans, re.IGNORECASE):
        print(phone_numbers)
        for el in phone_numbers:
            file.write(el)
            file.write("\n\n")
        sys.stdout.flush()
for element in name_list:
    if re.search(element, ans, re.IGNORECASE):
        print(fullname)
        file.write(fullname)
        file.write("\n")
        sys.stdout.flush()
```

```
personal_info_list = ["info", "about", "from"]
phone_info_list = ["phone", "number", "contact"]
name_list = ["name", "full name", "last name"]
committee_list = ["committee"]
service_list = ["service", "record", "long", "served"]
address_list = ["live", "where", "address"]
quit_cases = ["q", "quit", "exit"]
all_list = ["everything", "all"]
```