

How to Find All/Broken links using Selenium Webdriver

Deepak Varma <deepak.varma@citiustech.com>

Tue 10/22/2019 7:29 PM

To: Deepak Varma <deepak.varma@citiustech.com>

 1 attachments (60 KB)

Check broken links.txt;

What are Broken Links?

Broken links are links or URLs that are not reachable. They may be down or not functioning due to some server error

An URL will always have a status with 2xx which is valid. There are different HTTP status codes which are having different purposes. For an invalid request, HTTP status is 4xx and 5xx.

How to check the Broken Links and images

For checking the broken links, you will need to do the following steps.

1. Collect all the links in the web page based on <a> tag.
2. Send HTTP request for the link and read HTTP response code.
3. Find out whether the link is valid or broken based on HTTP response code.
4. Repeat this for all the links captured.

```
List<WebElement> links = driver.findElements(By.tagName("a"));

public void checkBrokenLinks(List<WebElement> links) {

    for(WebElement link : links) {

        url = link.getAttribute("href");

        if(url == null || url.isEmpty()){
            System.out.println("URL is either not configured for anchor tag or it is empty");
            //Assert.assertFalse(url == null || url.isEmpty(), "URL not configured or is empty");
            continue;
        }

        if(!url.startsWith("https://www.google.com")){
            System.out.println("URL belongs to another domain, skipping it.");
            continue;
        }

        //URLConnection class has methods to send HTTP request and capture HTTP response code.
        //So, output of openConnection() method (URLConnection) is type casted to HttpURLConnection.

        huc = (HttpURLConnection)(new URL(url).openConnection());

        huc.setRequestMethod("HEAD"); //We can set Request type as "HEAD" instead of "GET".
                                     //So that only headers are returned and not document body.

        huc.connect(); //On invoking connect() method, actual connection to url is established and the request is sent.

        respCode = huc.getResponseCode();

        //Assert.assertTrue(!(respCode >= 400), "URL is a broken link");
        if(respCode >= 400)
            System.out.println(url+" is a broken link");

        else
            System.out.println(url+" is a valid link");

    }

}
```