BFS Crawling : In BFS approach we will traverse all the URLs at the same depth first and then move to the URLs of next depth. As shown in the image above I have a set of web pages linked with each other. In BFS traversal, we will start with the seed URL. On seed page we have three links a, x and p. a is on top so we will first crawl to a URL and then x and then p URL. Once we are done with p, we will traverse b and y (depth 3) so the crawling sequence will be as follows:

       Seed,   a, x, p,   b, y, c, z, w

DFS Crawling: In depth first approach, we will go to a maximum depth of the top URL of the seed page first and then traverse to the second/next URL of seed page, So crawling sequence will be as follows:

       Seed (depth = 1), a (depth = 2), b (depth = 3), c (depth = 4), x (d = 5), p(d =1)

Comparison:

I am getting less no URLs in DFS approach than BFS approach. As I have shown earlier, URL x is on seed page so it should ideally have depth 0 but as we are using DFS approach we go in deep of URL a and at last depth we found x with depth = 5. So we cannot go ahead with x URL anymore and by this way we end up with less results.

By seeing first five URLs we can say DFS goes deep and deep of the first link whereas BFS stays to the main page only. DFS can be useful when we want to crawl some specific data from one page and other related data from few other pages.