

BUGS 3.txt

PREFACE

This file contains all of the known bugs inside of crawler .java. many of the bugs documented here will be easily seen when the program is run but it would be a good practice to document these kinds of things now instead of having them be found at runtime

1: output from commandline args.

This bug is basically due to the fact that now while running my program I have found that there is a fairly serious bug with my BFS/DFS iterators, I don't know how to effectively use them for one and also each time i tried to call them I got back an exception saying that i had a graph structure exception and that i could not run my iterator. I also ran into a problem while trying to do output on my edges and my nodes. I would need to add additional functionality to my crawlstate to be able to do these outputs effectively.

2: checking for text/html

In my Crawler program I have a check that looks for a text/html file type and I found that in testing my program over multiple webpages, some webpages that I know are html webpages were returned to me as not being such. This bug I don't know how to fix in all honesty and would take serious debugging to figure out.

3. canonicalizing urls

This is a major issue with my program I am trying to use regex to strip off the ending extension and then add the relative url to the current url but sometimes I run into errors where it just doesn't work. If you want to try a url that I am sure works through, run the program on www.fseven.org it is my personal blog website and I know it works :)

4. robots.txt

This may or may not be a bug in my programming. basically I am checking for the existence of a robots meta tag using the robot handler class given by professor lane and I think my method that implements this function works but I am not 100% sure. SO I am marking it as a bug anyways.

5. bugs.txt

This bugs.txt file may be the most poorly written bugs documentation ever. as I am writing this I am laying on my bed falling asleep and typing with my eyes closed. This is probably not the most tactful way to approach writing this kind of file but I hope that by some miraculous flap of a butterfly's wings that the bits inside my computer's registers will flip in

just the right way such that it produces a completely legible and fantastically well documented bugs.txt file/ Crawler program. In the event that this comes to pass, I will jump up and down for joy for 3 hours straight.

-Nialls Chavez