**GCrawlers: Christopher Beall, Helen Jiang, Brian Metzger** 

# **Project Demonstration**

#### **User instructions:**

## Prerequisites:

- 1) Internet connection
- 2) Desktop web browser
  - a) Web page minimum size of 900px by 1000px
  - b) Known supported web browser and versions:
    - i) Google Chrome version 78 and up
    - ii) Safari version 13.0 and up
    - III) Firefox version 70 and up

### Steps:

- 1) Navigate to <a href="https://gcrawler-test.herokuapp.com/search">https://gcrawler-test.herokuapp.com/search</a> using one of the web browsers stated above.
- 2) Fill out the form to begin a search:
  - a) Full Starting link: This is the full starting link to begin the web search. This field is required.
  - b) Choose a keyword: This is the word to stop the crawler when the word has been encountered on the page. This field is optional.
  - c) Choose a search type (required):
    - i) Choose Depth First Search to have the crawler randomly return a link on each web page
    - ii) Choose Breadth First Search to have the crawler return every link on each web page
  - d) Page limit (required):
    - i) If Depth First Search selected: choose a page limit between 1 and 10
    - ii) If Breadth First Search selected: choose a page limit between 1 and 3
  - e) Click "Search" to begin crawling.

#### Results page:

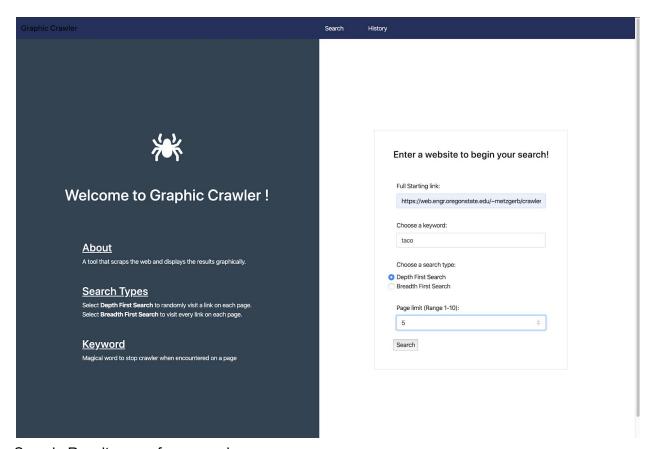
- a) The tree for your results is displayed here. The instructions for your results page can be found on the top right corner by hovering over the "Instructions" link.
  - i) Each node is a link returned from the search with the titles of each web page displayed on the node
  - ii) Each node connected by a line has a parent-child relationship
  - iii) Click on each node to be directed to the link through a new tab or window

- iv) Pan to move the tree around or scroll to zoom in and out
- v) Color scheme:
  - (1) Green: web pages with status < 400
  - (2) Red: web pages with status >= 400
  - (3) Yellow: web page containing the keyword from search form
- 4) Click on the History tab to view browser's history:
  - a) The history tab contains all of the previous searches saved on the browser
    - i) Each row contains the date of the search, the starting link, the search type, keyword if any, and the max page length
  - b) Hit the search button to search again using the criteria displayed on the table row
    - i) User will be automatically directed to the results page
  - c) Searches made from the history tab is not re-saved on the table to avoid duplicate entries
  - d) Clearing browser history will clear the history table
  - e) Closing out of the tab and returning to the history page will not clear the history table
  - f) Searches made in incognito mode will not be saved to the history table once the tab is closed

#### Search example:

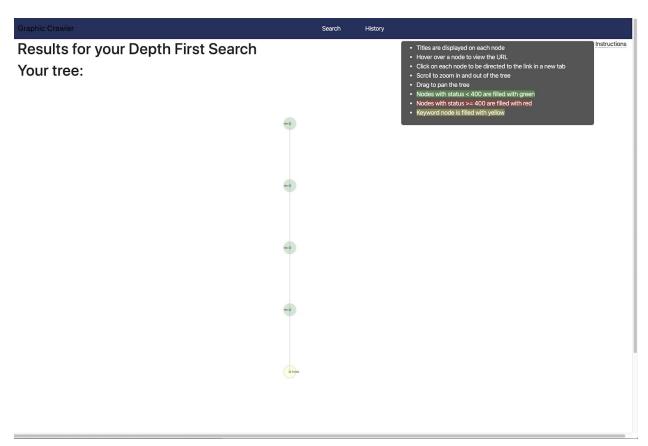
Use the following criteria for the search form:

- 1) Full Starting link: <a href="https://web.engr.oregonstate.edu/~metzgerb/crawler/root.html">https://web.engr.oregonstate.edu/~metzgerb/crawler/root.html</a>
- 2) Choose a keyword: taco
- 3) Choose a search type: Depth First Search
- 4) Page limit: 5



## Sample Results page from search:

- 1) Drag or zoom the tree to bring the full tree into view
- 2) There are 5 nodes; 4 green nodes and 1 yellow node.
  - a) Click on each green node to be directed to:
    - i) <a href="https://web.engr.oregonstate.edu/~metzgerb/crawler/root.html">https://web.engr.oregonstate.edu/~metzgerb/crawler/root.html</a>
    - ii) https://web.engr.oregonstate.edu/~metzgerb/crawler/2depth-1.html
    - iii) https://web.engr.oregonstate.edu/~metzgerb/crawler/relative.html
  - b) Click on the yellow node to be directed to web page containing the keyword "taco":
    - i) <a href="https://web.engr.oregonstate.edu/~metzgerb/crawler/index.html">https://web.engr.oregonstate.edu/~metzgerb/crawler/index.html</a>
  - c) Please note that your results page might be different since the Depth First Search is visiting a <u>random</u> link on each page



## History tab:

1) Click this history tab to see that your search has been added to the last row of the table

