

Class Notes

<p>Topic Goals/Objectives Essential Question</p>	<p>What is a web crawler? C.2.4 Describe how a web crawler function</p>
<p>What is a web robot?</p> <p>How do web crawlers work?</p> <p>What can a robots.txt file control?</p>	<p>Concept/Examples What is a web crawler?</p> <p>Who uses web crawlers?</p> <p>How can a web site control how a search engine accesses the content? Using a robots.txt file.</p> <p>What can a robots.txt file control?</p> <p>What are some of the command line options for using wget?</p>
	<p>Practice Problems</p> <p>Retrieve files using a web crawler utility named wget</p> <ul style="list-style-type: none"> • Connect to your Cloud9 java workspace • from a terminal window: <ul style="list-style-type: none"> ▪ cd lessons ▪ issue the wget command: wget hwmath.net/IBCS/files/StateNode.java ▪ issue the wget command: wget hwmath.net/IBCS/files/States.java ▪ Verify that you have downloaded two files using the ls command: