

Web Intelligence

EXERCISE FOR LECTURE 2

1. Implement a function, which will parse a robots.txt file and return any access restrictions that may apply to your crawler.
2. Find 5-10 robots.txt file on the web and test your parser on these.

The robots.txt parser will come in handy for the assignment. It is very important to comply with restrictions in such a file. As soon as you are done with the parser, you may start working on the crawler in your assignment !