NWebCrawler: A Stupid C# Web Crawler

# Overview

Web crawler is a program that exploits the graph structure of the Web, move from page to page, and retrieve web pages and add them to a local repository.

# Design

NWebCrawler V 1.0.0 implements a basic multi-threaded crawler. The crawler maintains a list of unvisited URLs called the frontier. The list is initialized with seed URLs provided by a user. Each crawling loop involves picking the next URL to crawl from the frontier, parsing the retrieved page to extract the URLs, and finally adding the unvisited URLs to the frontier.



# Reference

1. “*Crawling the Web*”. <http://pdf.aminer.org/000/003/078/crawling_the_web.pdf>.