NWebCrawler: A Stupid C# Web Crawler

# Overview

实现NWebCrawler V1.0.2.



# Design

1.0.2解决老版本（V1.0.0）的三个问题：

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# 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>.