



# Documentul de testare al aplicației

Decembrie, 2020  
Versiune 1.0

## Web Crawler

**Academia Tehnica Militara “Ferdinand I”  
Realizat de:**

**Sd.sg.maj. Andrei Claudia  
Sd.sg.maj. Chiriță Gabriela  
Sd.sg.maj. Guțu Bogdan  
Sd.sg.maj. Manea Sebastian  
Sd.sg.maj. Mercheș Diana**

## 1. Introducere

Pentru a se efectua procesul de testare a aplicației, va fi nevoie de un calculator având sistem de operare Windows și o bună conexiune la Internet

Pe parcursul evaluării aplicației, se vor avea în vedere testele aflate în “Anexa #1”. Rezultatele vor fi trecute în “Anexa #2”, respectându-se convențiile stabilite.

### 1.1. Funcționalitățile testate

Pe parcursul testării, se va ține cont de un tabel cu funcționalități, prezent în “Anexa #1”, atașată documentului, și ce are următoarele coloane:

- ID: identificator unic al testului;
- Utilizator: tipul de utilizator care are acces la funcționalitate;
- Funcționalitate: prezentarea succintă a caracteristicii testate;
- Metodologia testării: modul în care se va executa testul;
- Posibile erori: erori ce pot apărea pe parcursul efectuării testului;

### 1.2. Criterii de evaluare

În procesul de testare, se vor folosi trei criterii de evaluare a testelor:

- acceptat: aplicația se comportă favorabil, în cadrul a trei teste consecutive, de același tip;
- respins: aplicația prezintă erori sau nu se comporta favorabil, în cadrul a trei teste consecutive, de același tip;
- amânat: aplicația trece printr-o serie imprevizibilă de răspunsuri favorabile, erori sau comportament nefavorabil, în cadrul a cinci teste consecutive, de același tip.

## Anexa #1: Funcționalități testate

ID	UTILIZATOR	FUNCȚIONALITATE	METODOLOGIA TESTĂRII	POSIBILE ERORI
T1	Client	Validare fișier de configurare	Se introduce fișierul de configurare de către client. Se realizează teste asupra fișierului de configurare pentru a vedea dacă datele sunt valide.	<ul style="list-style-type: none"> <li>Fișier de input inexistent.</li> <li>Datele de input sunt invalide.</li> </ul>
T2	Client	Validare URL	Clientul introduce de la tastatură fișierul cu URL-uri și se realizează validarea acestora.	<ul style="list-style-type: none"> <li>URL-urile sunt invalide.</li> </ul>
T3	Client	Obținere informații pe baza URL-urilor	Pe baza fișierului cu URL-uri introdus de client se realizează descărcarea locală.	
T4	Client	Web Crawler funcțional	Clientul introduce fișierul de configurare și așteaptă validarea acestuia. Pe baza URL-urilor, WebCrawler-ul realizează descărcarea, filtrarea, analizarea datelor și site map-ul.	<ul style="list-style-type: none"> <li>Fișier de input inexistent.</li> <li>Date invalide.</li> <li>URL greșit.</li> </ul>

## Anexa #2: Rezultate ale testării

ID	T1	T2	T3	T4
1	✓	✓	✓	✓
2	✓	✓	✓	✓
3	✓	✓	✓	✓
4	✓	✓	✓	✓

**Pentru T1:** Am introdus 4 fișiere diferite de configurare în urma cărora aplicația s-a comportat corespunzător. Pentru verificare, am afișat datele disponibile din fișiere pe ecran.

```

WELCOME TO WEBCRAWLER-MTA
In order to use the application you have the following possibilities:
[1]-->Crawling configuration file:crawl config.cnf
[2]-->Crawling sites from a given input file:crawl input.txt
[3]-->Crawling sites from a given input file with a specified type:crawl input.txt type png
[4]-->Filter downloaded sites by type:filter type png
[5]-->Filter downloaded sites by date:filter date zz/mm/yyyy
[6]-->Filter downloaded sites by dimension:filter size 150 kb
[7]-->Create sitemap in an output file:sitemap sitemap.txt
[8]-->Search word in all the downloaded files:search word
[9]-->Exit:exit
crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\config.cnf
Starting crawling config file..
Config file crawled.
Number of thread:50
Delay is: 100ms
Directory Path is: C:/Users/Diana/Desktop/CrawlerDownload
Log level is: 3
Depth is: 3
Enter new command

```

```

crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\config1.cnf
Starting crawling config file..
Config file crawled.
Number of thread:8
Delay is: 200ms
Directory Path is: C:/Users/Diana/Desktop/Crawler
Log level is: 5
Depth is: 5
Enter new command
crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\config2.cnf
Starting crawling config file..
Config file crawled.
Number of thread:4
Delay is: 5ms
Directory Path is: C:/Users/Diana/Desktop/Crawler1
Log level is: 2
Depth is: 2
Enter new command
crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\config3.cnf
Starting crawling config file..
Config file crawled.
Number of thread:300
Delay is: 300ms
Directory Path is: C:/Users/Diana/Desktop/CrawlerMTA

```

**Pentru T2:** Am introdus 4 fișiere diferite de input care conțin diferite URL-uri. Pentru verificare, am afișat listele de URL-uri pe ecran.

```
crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\config.cnf
Starting crawling config file..
Enter new command
crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\input.txt
Starting crawling input file..
URL list is: [https://mta.ro]
```

```
crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\config2.cnf
Starting crawling config file..
Enter new command
crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\input2.txt
Starting crawling input file..
URL list is: [https://mta.ro/, https://www.netflix.com, https://wiki.mta.ro/]
```

```
crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\config3.cnf
Starting crawling config file..
Enter new command
crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\input3.txt
Starting crawling input file..
URL list is: [https://mta.ro, https://www.netflix.com, https://wiki.mta.ro/, https://www.wikipedia.org/]
```

```
crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\config4.cnf
Starting crawling config file..
Enter new command
crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\input4.txt
Starting crawling input file..
URL list is: [https://www.netflix.com, https://www.wikipedia.org/]
```

**Pentru T3:**

Descărcare tuturor fișierelor local, în directorul specificat.

```
Saving fileC:\Users\Diana\Desktop\CrawlerMTA\mta.ro/wp-content/uploads/2019/09/cropped-android-chrome-512x512-192x192.png
Saving fileC:\Users\Diana\Desktop\CrawlerMTA\mta.ro/wp-content/uploads/2019/09/cropped-android-chrome-512x512-180x180.png
Saving fileC:\Users\Diana\Desktop\CrawlerMTA\mta.ro/about-us
Saving fileC:\Users\Diana\Desktop\CrawlerMTA\mta.ro/legislatie
Saving fileC:\Users\Diana\Desktop\CrawlerMTA\mta.ro/conducere
Saving fileC:\Users\Diana\Desktop\CrawlerMTA\mta.ro/regulamente-de-organizare-si-functionare
Saving fileC:\Users\Diana\Desktop\CrawlerMTA\mta.ro/wp-content/uploads/2020/03/organigrama.pdf
Saving fileC:\Users\Diana\Desktop\CrawlerMTA\mta.ro/posturi-vacante
Saving fileC:\Users\Diana\Desktop\CrawlerMTA\mta.ro/programe-si-strategii
Saving fileC:\Users\Diana\Desktop\CrawlerMTA\mta.ro/capante-privind-starea-institutiei
```

În fișierul de configurare am setat tipul pentru descărcare „png”. Astfel, observăm o descărcare doar a fișierelor de tip png.

```
crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\config.cnf
Starting crawling config file..
Enter new command
crawl C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\input.txt
Starting crawling input file..
Enter new command
Saving fileC:\Users\Diana\Desktop\DownloadCrawler\mta.ro/wp-content/uploads/2019/09/cropped-android-chrome-512x512-32x32.png
Saving fileC:\Users\Diana\Desktop\DownloadCrawler\mta.ro/wp-content/uploads/2019/09/cropped-android-chrome-512x512-192x192.png
Saving fileC:\Users\Diana\Desktop\DownloadCrawler\mta.ro/wp-content/uploads/2019/09/cropped-android-chrome-512x512-180x180.png
```

În fișierul de configurare am setat tipul pentru descarcare „html”. Astfel, observăm o descărcare doar a fișierelor de tip html.

```
C:\WWW C:\Users\Diana\Desktop\WebCrawlerMTA\WebCrawler\resources\input2.txt
Enter new command
https://wiki.mta.ro//robots.txt
https://codex.nflxext.com/%5E3.0.0/truthBundle/webui/1.22.5-shakti-js-v526798ba/js/js/bootstrap.js,common%7Cbootstrap.js/2/0b4003414b0647/
https://codex.nflxext.com/%5E3.0.0/truthBundle/webui/1.22.5-shakti-js-v526798ba/js/js/signup%7Chome%7CcourStory%7Cfuji%7CfujiFrameworkClic
https://codex.nflxext.com/%5E3.0.0/truthBundle/webui/1.22.5-shakti-css-v526798ba/css/css/less%7Ccore%7Cerror-page.less/1/DwHvP3NXLn8GM0/n
Saving fileC:/Users/Diana/Desktop/Crawler1/curl.haxx.se/docs/CVE-2020-8169.html
Saving fileC:/Users/Diana/Desktop/Crawler1/curl.haxx.se/docs/CVE-2020-8169.html
https://codex.nflxext.com/%5E3.0.0/truthBundle/webui/1.22.5-shakti-css-v526798ba/css/css/less%7Cpages%7Chome%7CcourStory%7Cfuji%7Cfuji_les
```

În fișierul de configurare am setat tipul pentru descarcare „docx”. Astfel, observăm o descărcare doar a fișierelor de tip docx.

```
https://www.mta.ro/international/eea-grant-ro/robots.txt
https://www.mta.ro/international/interuniversity-partners-ro/robots.txt
https://www.mta.ro/international/international-semester-ro/robots.txt
Saving fileC:/Users/Diana/Desktop/CrawlerMTA/www.mta.ro/international/wp-content/uploads/2020/08/Application_Form_ITS_MTAFI_2021.docx
https://www.mta.ro/international/conferences-ro/robots.txt
```

## Pentru T4:

Cu ajutorul fișierelor folosite anterior, am realizat diferite operații asupra fișierelor descărcate.

## Verificare cuvinte cheie:

```
search about
Starting searching word..
There are 3 files that contain the word [about]:
C:\Users\Diana\Desktop\Crawler1\curl.haxx.se\docs\CVE-2020-8169.html
C:\Users\Diana\Desktop\Crawler1\upt.ro\Universitatea-Politehnica-Timisoara_en.html
C:\Users\Diana\Desktop\Crawler1\www.mta.ro\geoint\index.html
```

## Creare Sitemap:

```
sitemap sitemap.txt
Starting creating sitemap..
Sitemap created.
```

```
sitemap.txt - Notepad
File Edit Format View Help
mta.ro/
  category/
  comments/
  evenimente/
  stiri-anunturi/
  wp-content/
    uploads/
      2019/
        08/
        09/
          cropped-android-chrome-512x512-180x180.png
          cropped-android-chrome-512x512-192x192.png
          cropped-android-chrome-512x512-32x32.png
      2020/
        03/
        11/
        12/
  wp-includes/
  wp-json/
  oembed/
```

**Filtrarea elementelor descărcate local** în funcție de mai multe criterii(data descărcării, tipul fișierelor, dimensiunea fișierelor):

```
Enter new command
Filter date 20/12/2020
Starting filtering by date..
Files downloaded before 20/12/2020 are: Application-form_Model.docx 15/12/2020
Files downloaded before 20/12/2020 are: Application_Form_ITS_MTAFI_2021.docx 15/12/2020
Enter new command
Filter type docx
Starting filtering by type..
List of the docx files in the specified directory:
C:\Users\Diana\Desktop\CrawlerMTA\www.mta.ro\international\wp-content\uploads\2019\03\Application-form_Model.docx
C:\Users\Diana\Desktop\CrawlerMTA\www.mta.ro\international\wp-content\uploads\2020\08\Application_Form_ITS_MTAFI_2021.docx
Enter new command
Filter size 150 kb
Starting filtering by size..
Files with dimension less than 150 KB are:
C:\Users\Diana\Desktop\CrawlerMTA\www.mta.ro\international\wp-content\uploads\2020\08\Application_Form_ITS_MTAFI_2021.docx 1 KB
```

Observații:

Coloanele din tabelul “Anexa #2”, corespunzătoare testelor, se vor completa cu simboluri specifice:

✓ : acceptat

✗ : respins

? : amânat

## 2. Anexa

### 2.1. Tehnologii folosite

JAVA

### 2.2. Acronime

SRS	Software Requirements Specification
URL	Uniform Resource Locator
CLI	Command Line Interface