

Documentare de testare al aplicației

**Web Crawler**

Academia Tehnică Militară „Ferdinand I”

Realizat de:

Sd. Sg. Maj. Albu Alexandru

Sd. Sg. Maj. Călugăru Teodor

Sd. Sg. Maj. Irina Mirea

Sd. Sg. Maj. Ghiță Cristian

1. Introducere

Pentru realizarea acestor teste asupra aplicației s-a utilizat un calculatore având sistem de operare Windows și o conexiune bună la Internet.

Se vor respecta testele aflate in „Anexa #1”. Rezultatele vor fi înregistrate în „Anexa #2”.

* 1. **Funcționalități testate**

Testele efectuate se vor afla în tabelul de funcționalități din „Anexa#1”:

* ID: identificatorul unic al testului
* Funcționalitate: se prezintă o funcționalitate specifică
* Comportament corespunzător: se descrie comportamentul pe care funcționalitatea trebuie să o aibă
* Test efectuat: se va descrie testul efectuat pentru testare
* Rezultat
  1. **Raționalizarea deciziei asupra testelor**

Pentru reprezentarea rezultatului testelor, se vor folosi următoarele criterii de evaluare.

* PASSED: testul a rulat cu succes, evidențiind funcționarea corespunzătoare a funcționalității în condițiile date de test
* FAILED: testul nu a rulat cu succes, evidențiind funcționarea defectoasă a funcționalității în condițiile testului

**Anexa #1: Funcționalități testate**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Funcționalitate | Comportament | Test efectuat | Rezultat |
| T1 | Realizarea crawl-ului | Se va face crawl pe diferite adrese URL a unor site-uri cu diferite topologii | Se vor lua adrese URL, se va verifica dacă aceasta este validă | PASSED |
| T2 | Realizarea sitemap-ului pe baza topologiilor reținute local | Se va realiza un sitemap pe baza topologiilor locale | Se va fișierul în care au fost reținute topologiile adreselor URL și se va realiza sitemap-ul acestora | PASSED |
| T3 | Filtrarea pe baza dimensiunii | Se vor șterge fișierele care vor depăși o anumită dimensiune dată ca argument | Se inserează fișiere de diferite dimensiuni în componența unor topologii, se va realiza filtrarea și se va observa comportamentul | FAILED |
| T4 | Căutarea pe baza unor cuvinte cheie | Se vor afișa în linie de comandă | Se vor da o serie de cuvinte cheie ca și argument, se vor găsi acestea în componența topologiilor și se va scrie într-un fișier | PASSED |
| T5 | Filtrarea pe baza tipului fișierului | Se vor șterge fișierele care nu au tipul corect | Se vor insera fișiere de diferite tipuri în componența unor topologii, se vor specifica tipul de fișiere permis și se vor șterge toate fișierele ce nu respectă regula | PASSED |

**T1. Realizarea crawl-ului**

S-a realizat crawl pe 5 adrese URL după care s-a observat dacă s-au realizat topologiile corespunzătoare, descărcând resursele disponibile pe un director local.

Input: <https://mta.ro/> Rezultat: PASSED

Input: <https://mapn.ro/> Rezultat: PASSED

Input: <https://emag.ro/> Rezultat: PASSED

Input: <https://www.armyacademy.ro/> Rezultat: PASSED

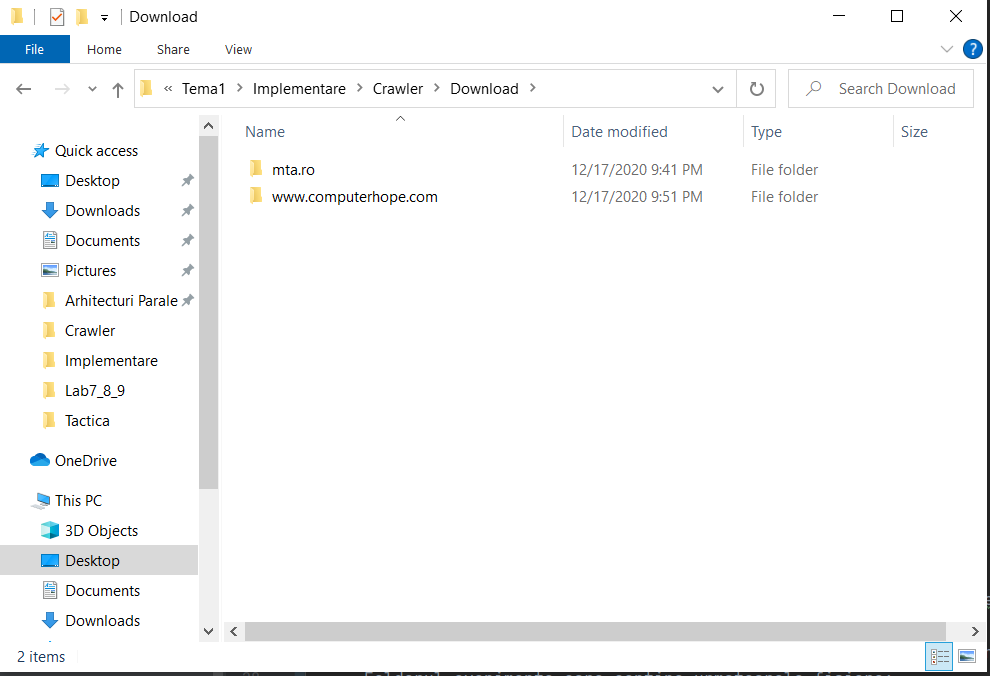
Input: <https://www.academiadepolitie.ro/> Rezultat: PASSED

**T2. Realizarea sitemap-ului pe baza topologiilor reținute local**

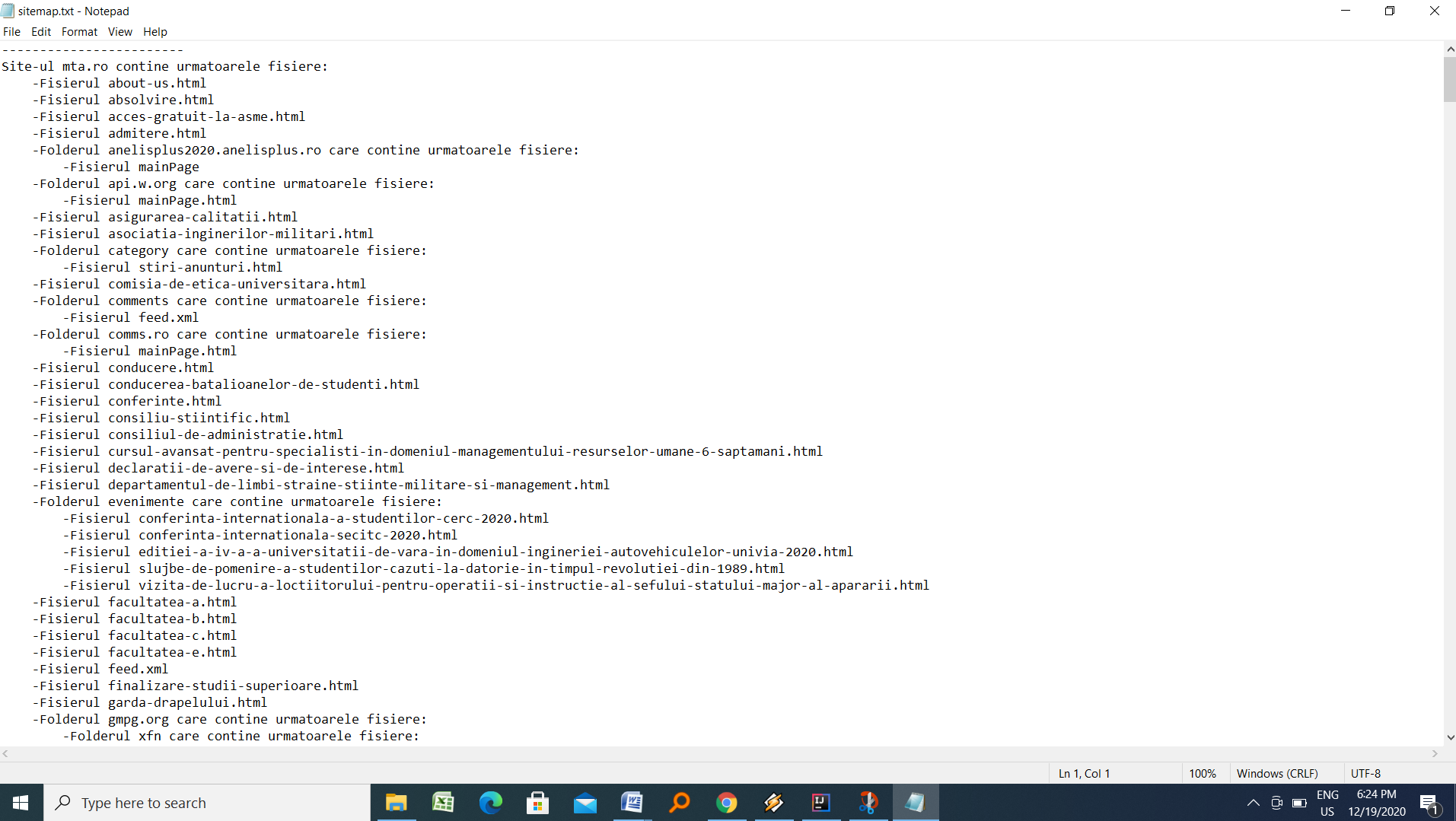
După ce avem descărcate toate resursele de pe site-urile din fișierul sites.txt în folderul /Download se va realiza un sitemap care va conține structura tuturor acestor site-uri (nu se va construi câte un sitemap pentru fiecare site în parte).

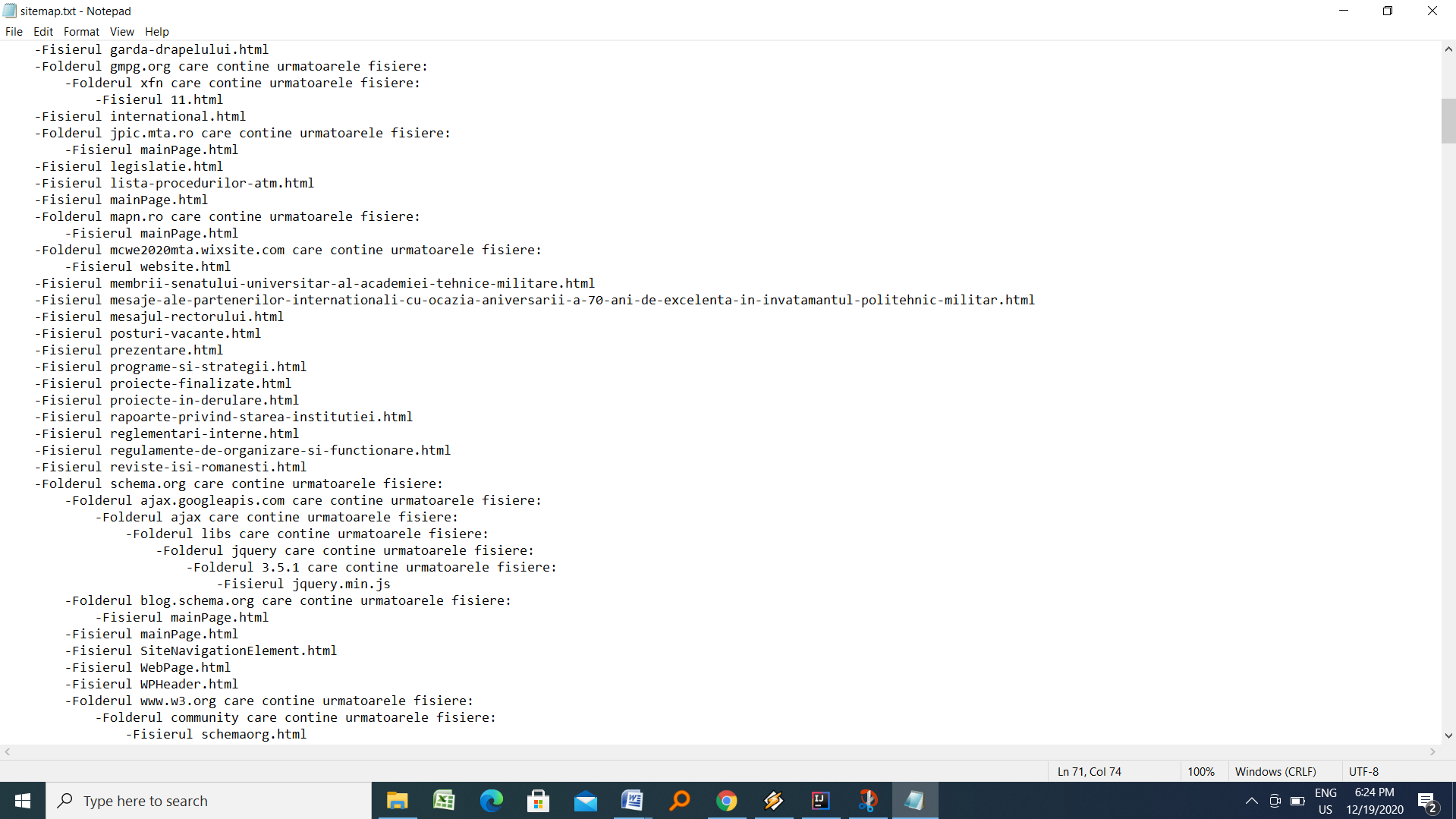
Vom verifica această funcționalitate pentru câteva configurații ale fișierului /Download:

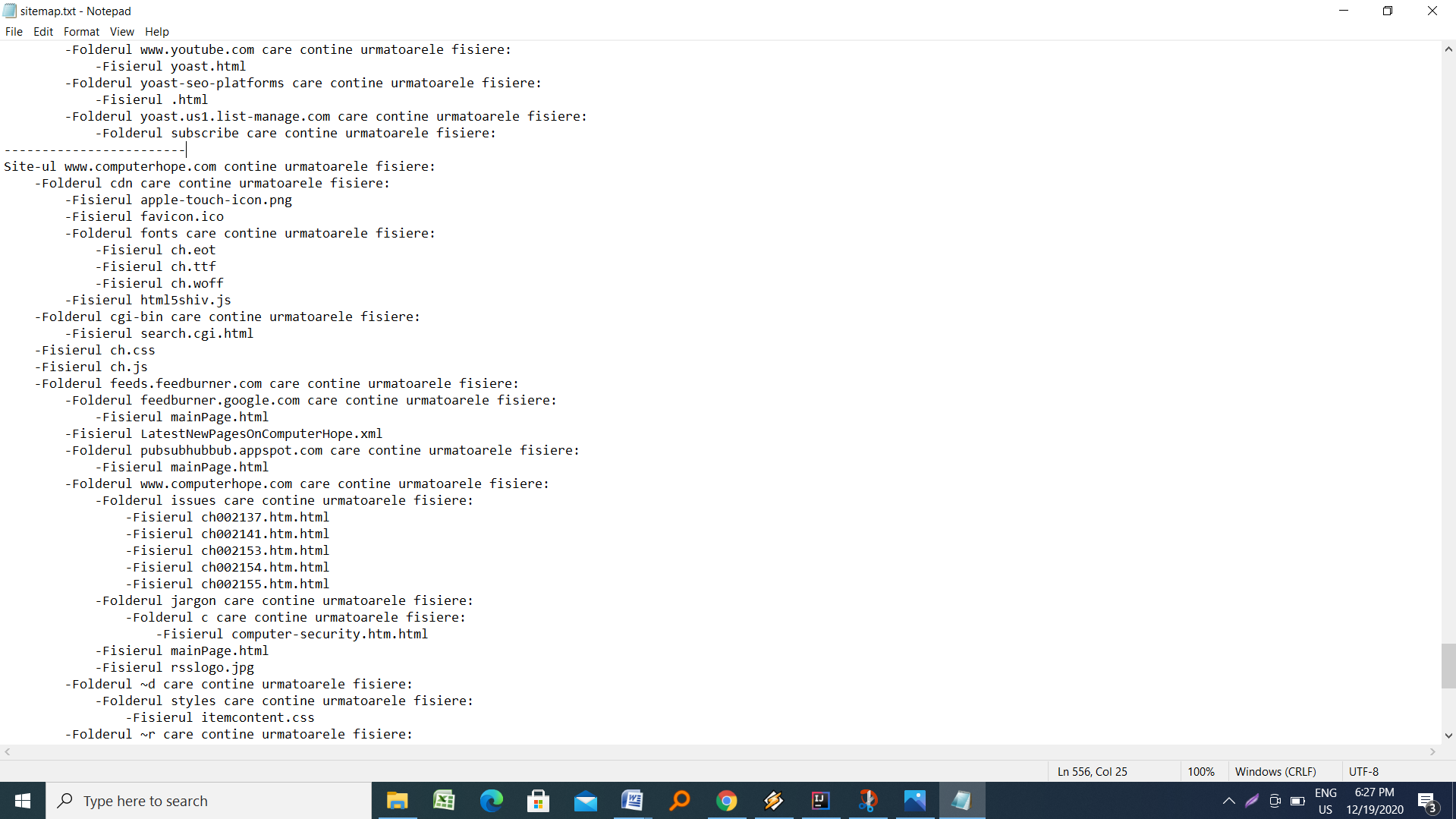
1. Input:



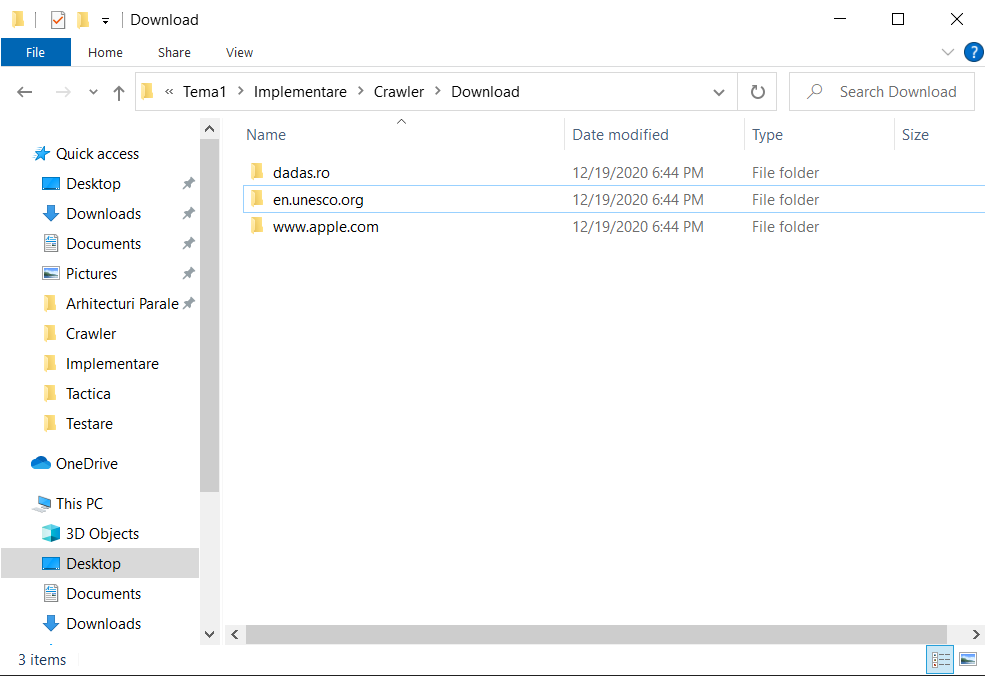
Output (fișierul sitemap):



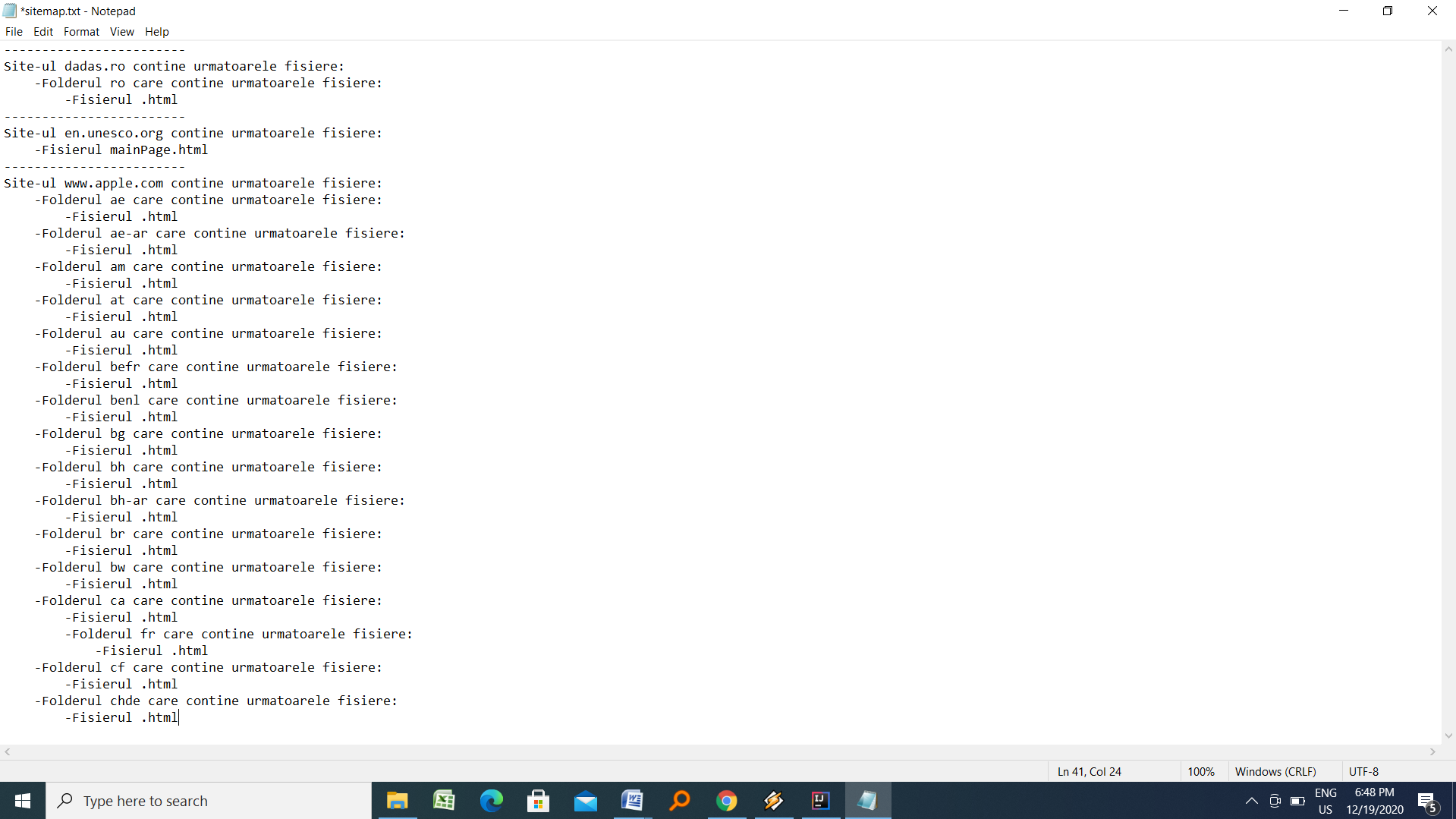




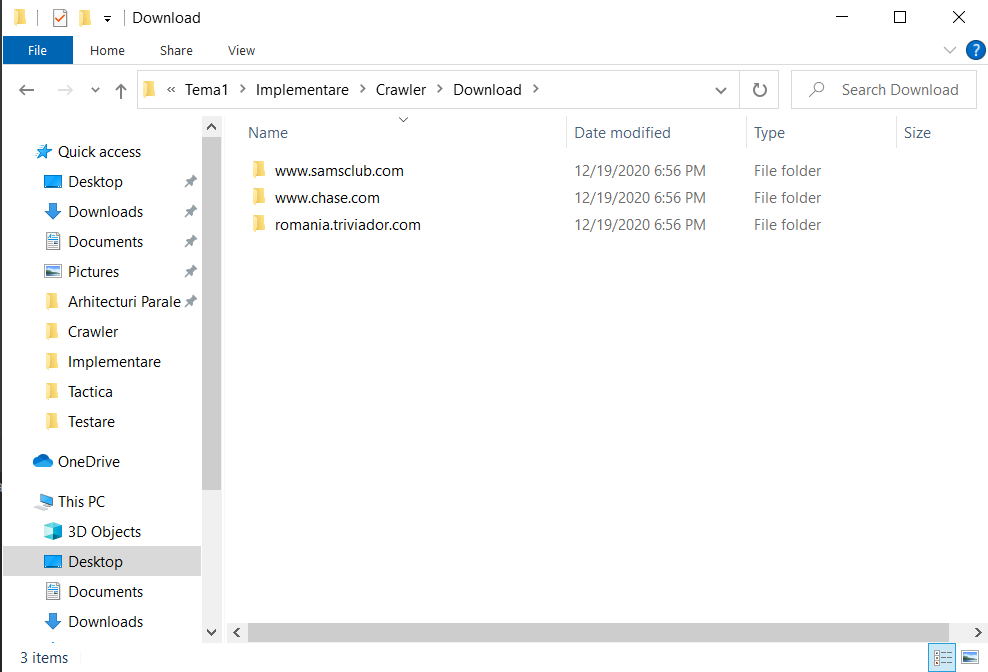
1. Input:



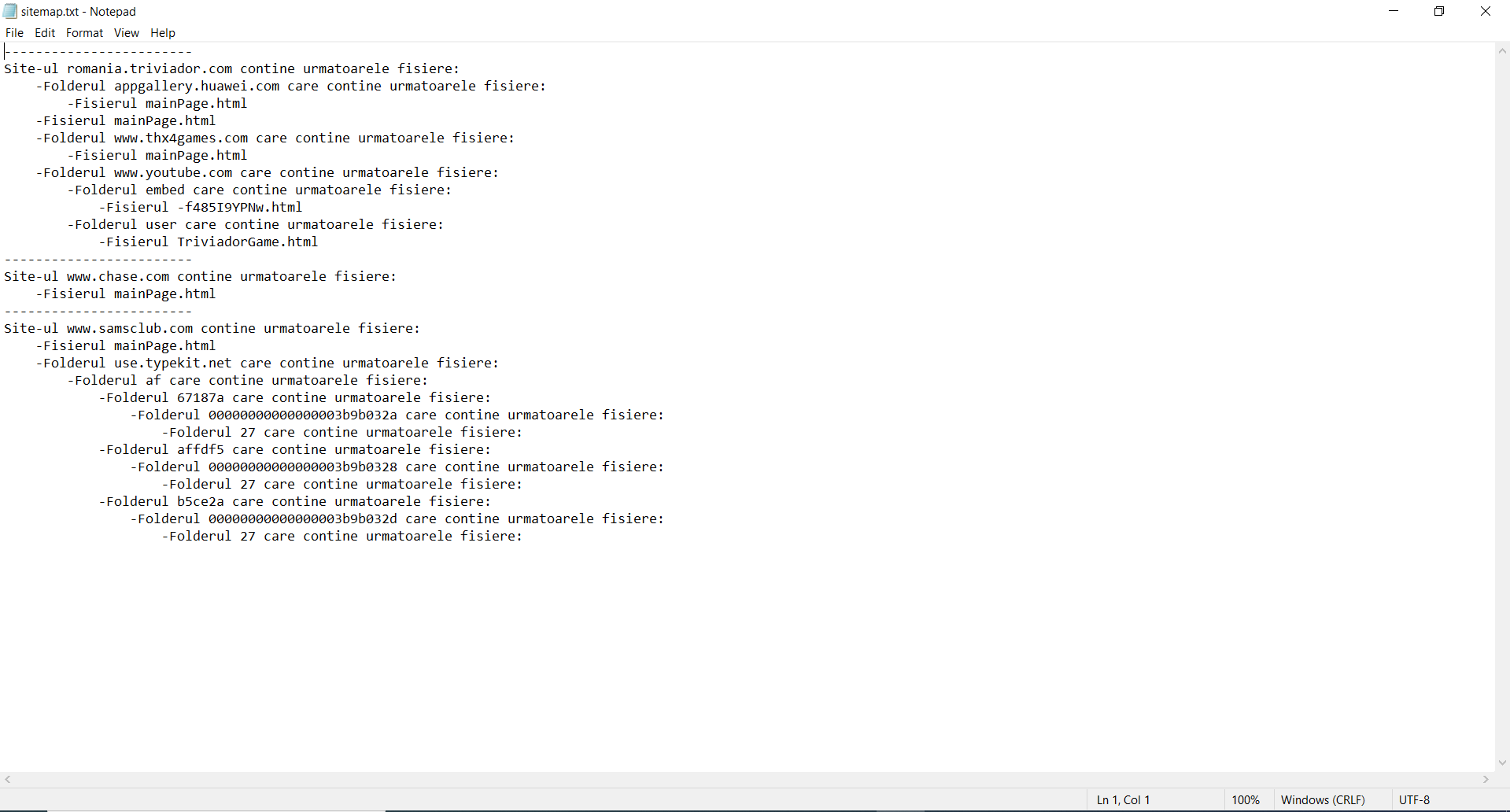
Output (fișierul sitemap):



1. Input:



Output:



**T3. Filtrarea pe baza dimensiunii**

După ce avem toate resursele descărcate prin intermediul crawler-ului se vor putea filtra în funcție de dimensiunea maximă a acestora.

S-au inserat fișiere de dimensiuni variate în topologii deja existente în scop de testare din moment ce într-un site nu există o multitudine de fișiere de dimensiuni mari și s-au efectuat asupra lor observații referitor la comportament:

Input: filtrare a fișierelor până în 2 MB

Output: Anumite fișiere peste 2kB nu au fost șterse(FAILED)

Input: filtrare a fișierelor până în 1 MB

Output: Fișierele au fost șterse cu succes(PASSED)

Input: filtrarea fișierelor până în 5MB

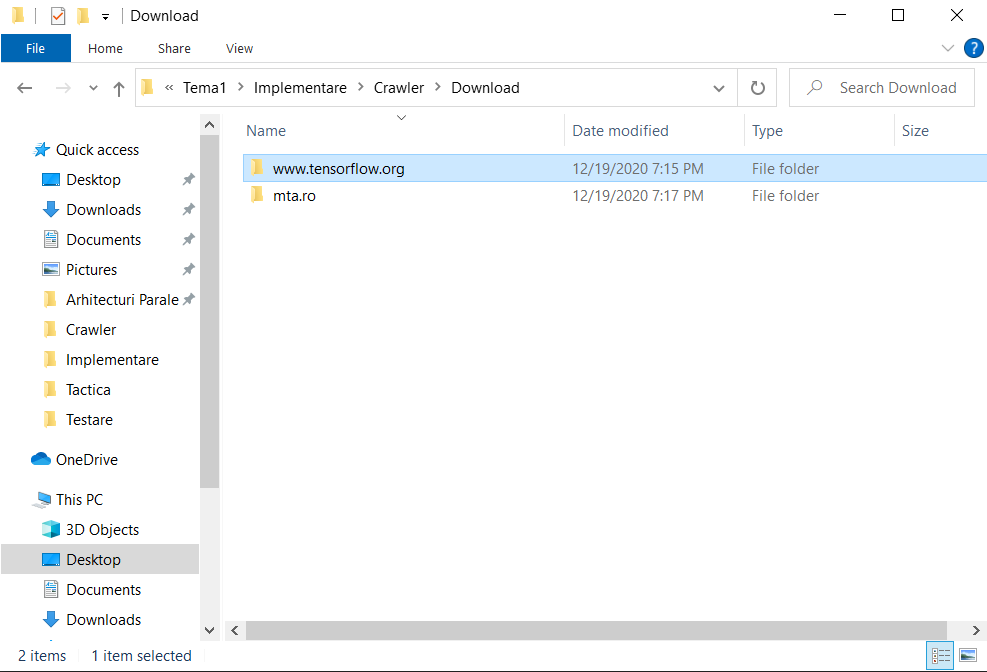
Output: niciun fișier nu a fost șters(fișiere peste 5MB fiind inexistente)(PASSED)

**T4. Căutarea pe baza unor cuvinte cheie**

După ce avem descărcate toate resursele de pe site-urile din fișierul sites.txt în folderul /Download utilizatorul va avea posibilitatea de a căuta anumite cuvinte cheie in fișierele descărcate(se va realiza o căutare a acestor cuvinte în toate fișierele descărcate).

Vom verifica această funcționalitate pentru câteva opțiuni ale utilziatorului.

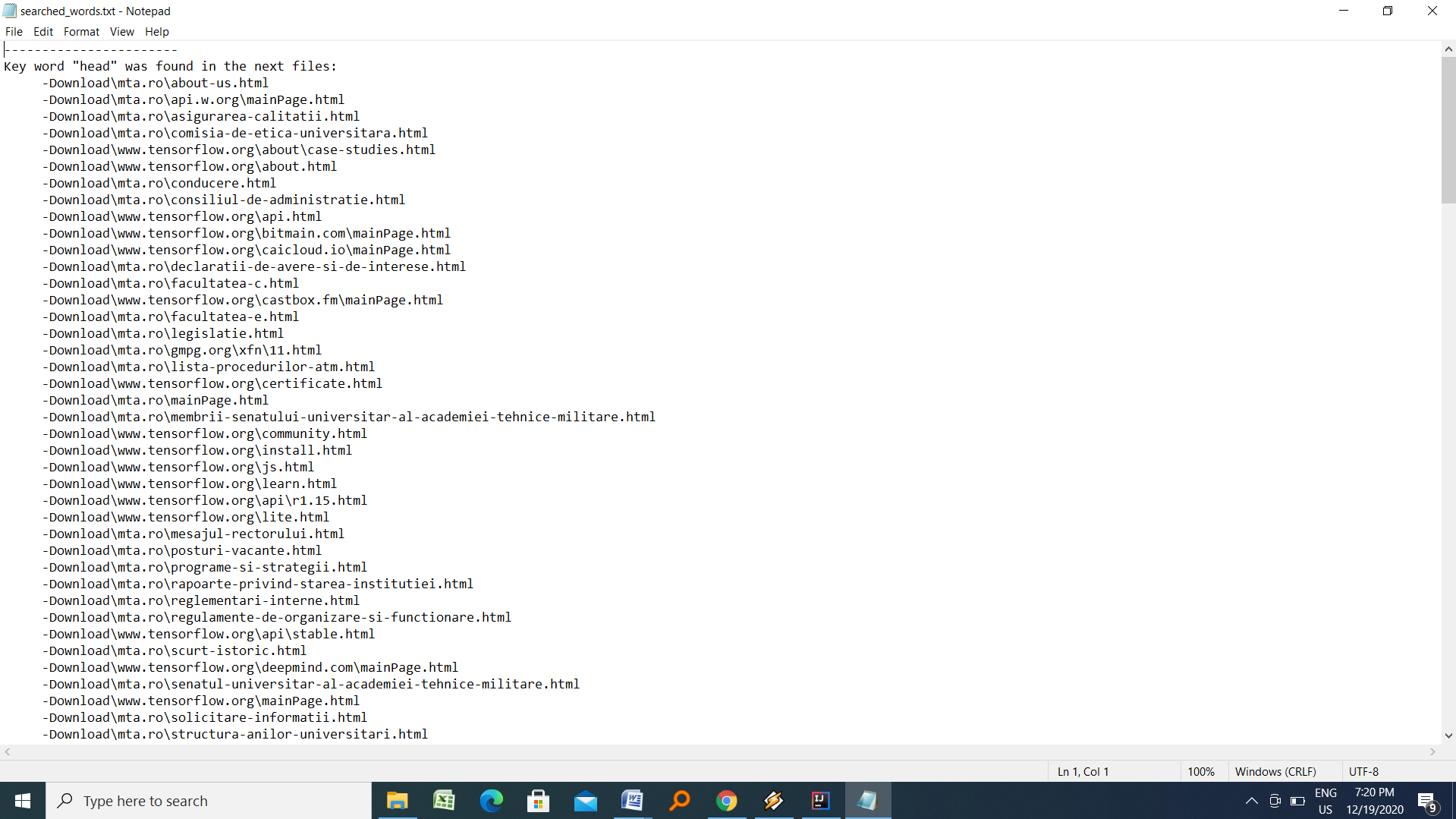
Am folosit următoarea structură pentru folderul /Download:

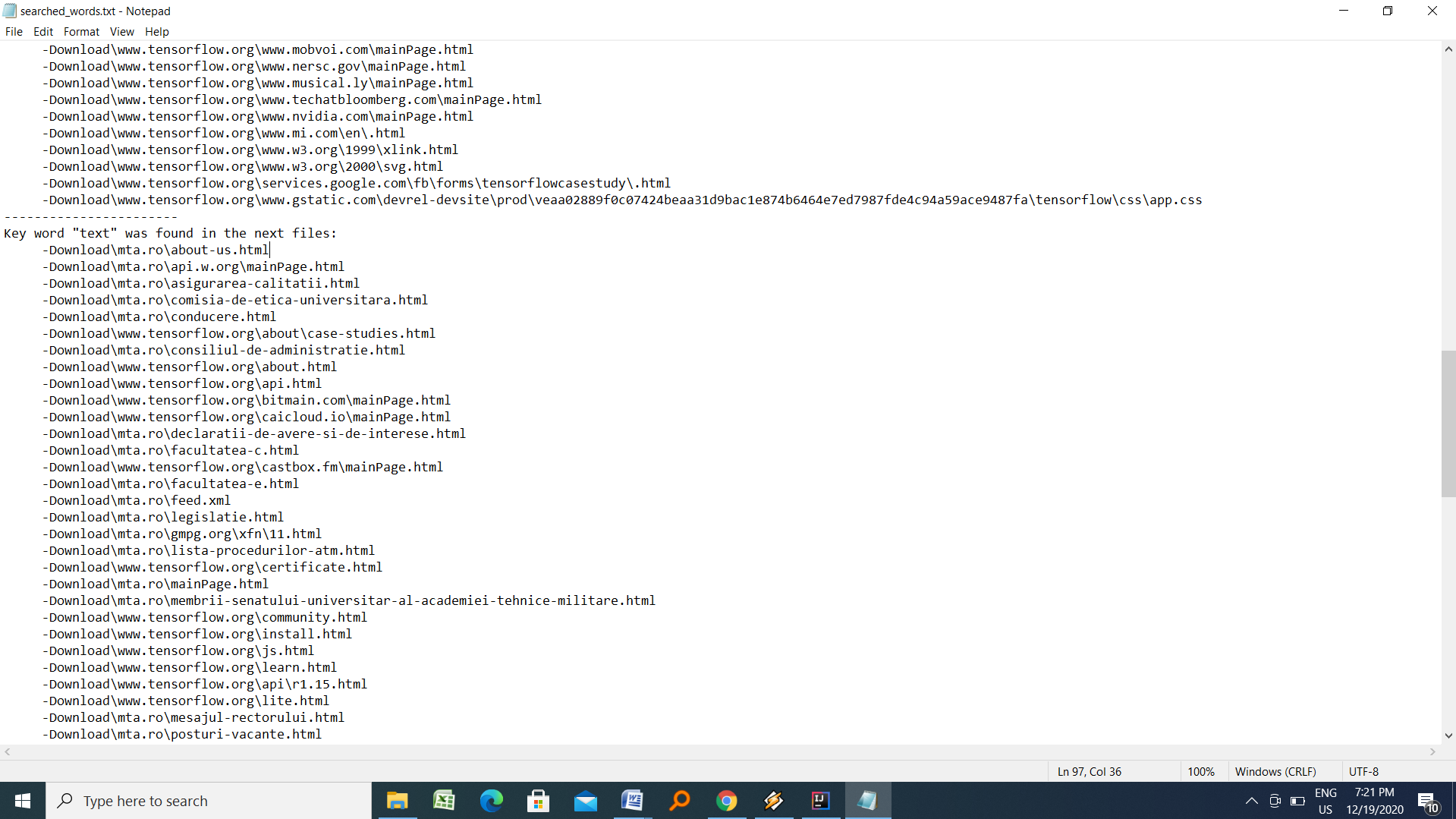


1. Input:

Cuvinte căutate: text

Output(fișierul searched\_words.txt):

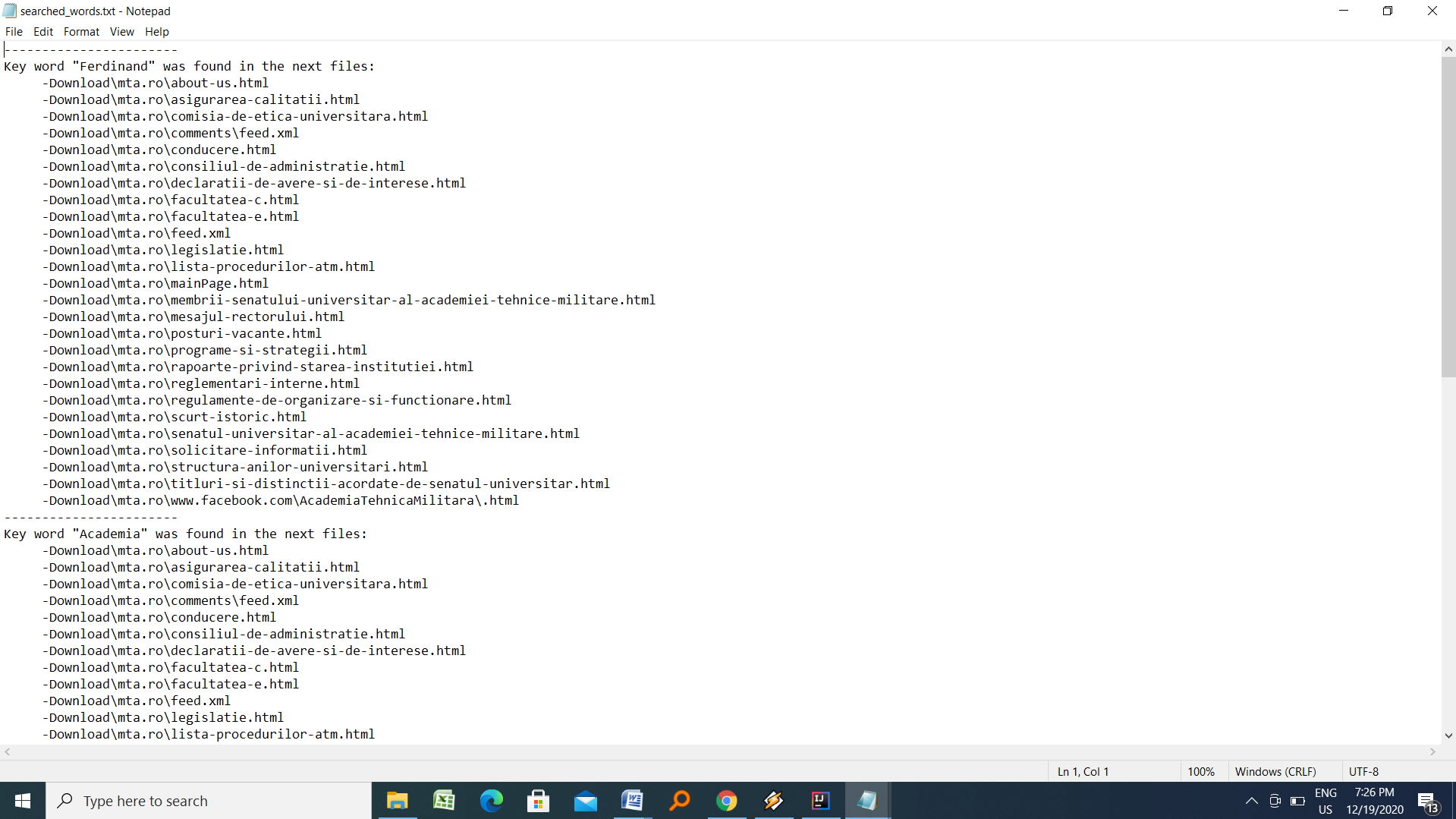


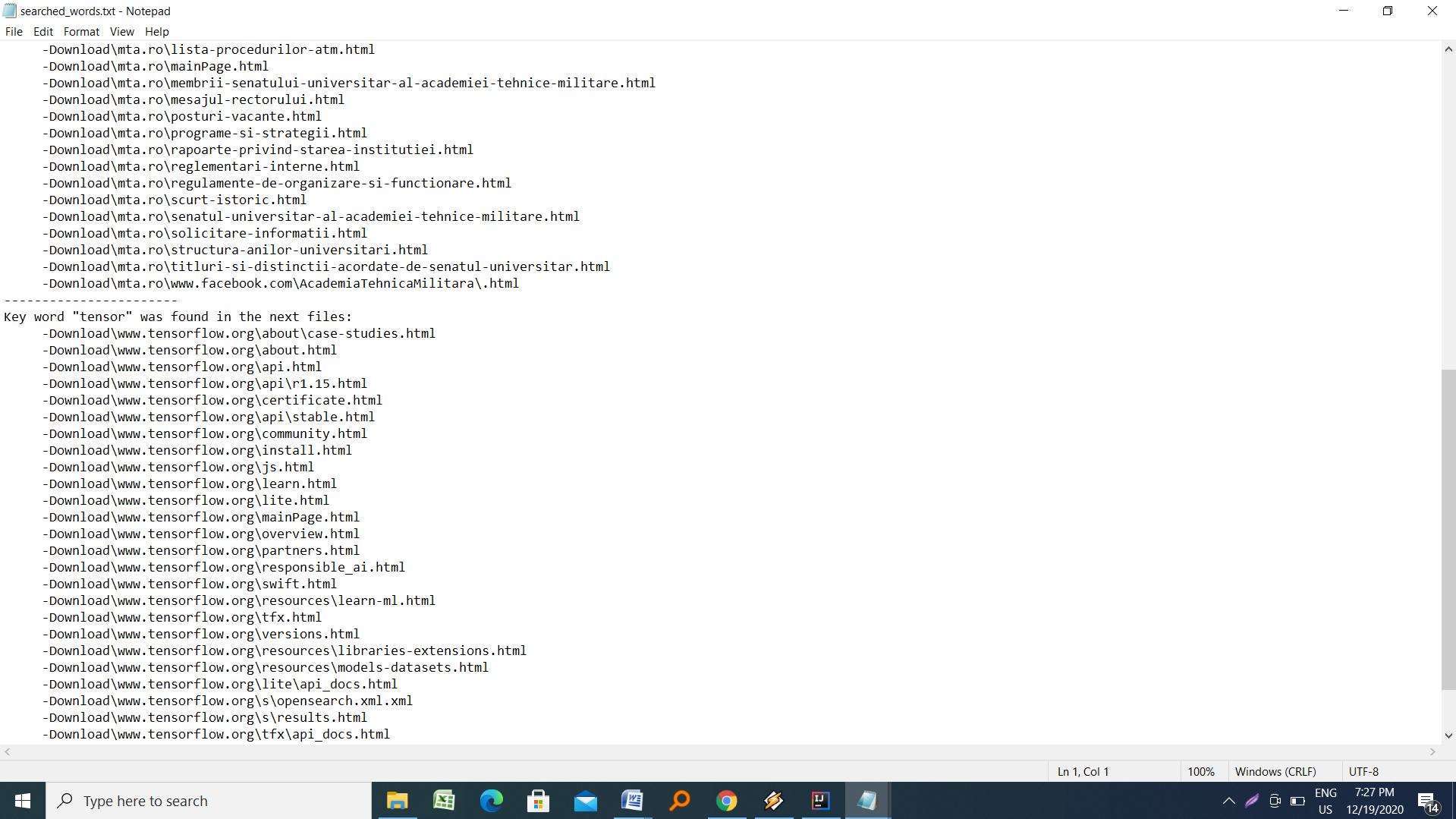


1. Input:

Cuvinte căutate: tensor, Academia, Ferdinand

Output(fișierul searched\_words.txt):

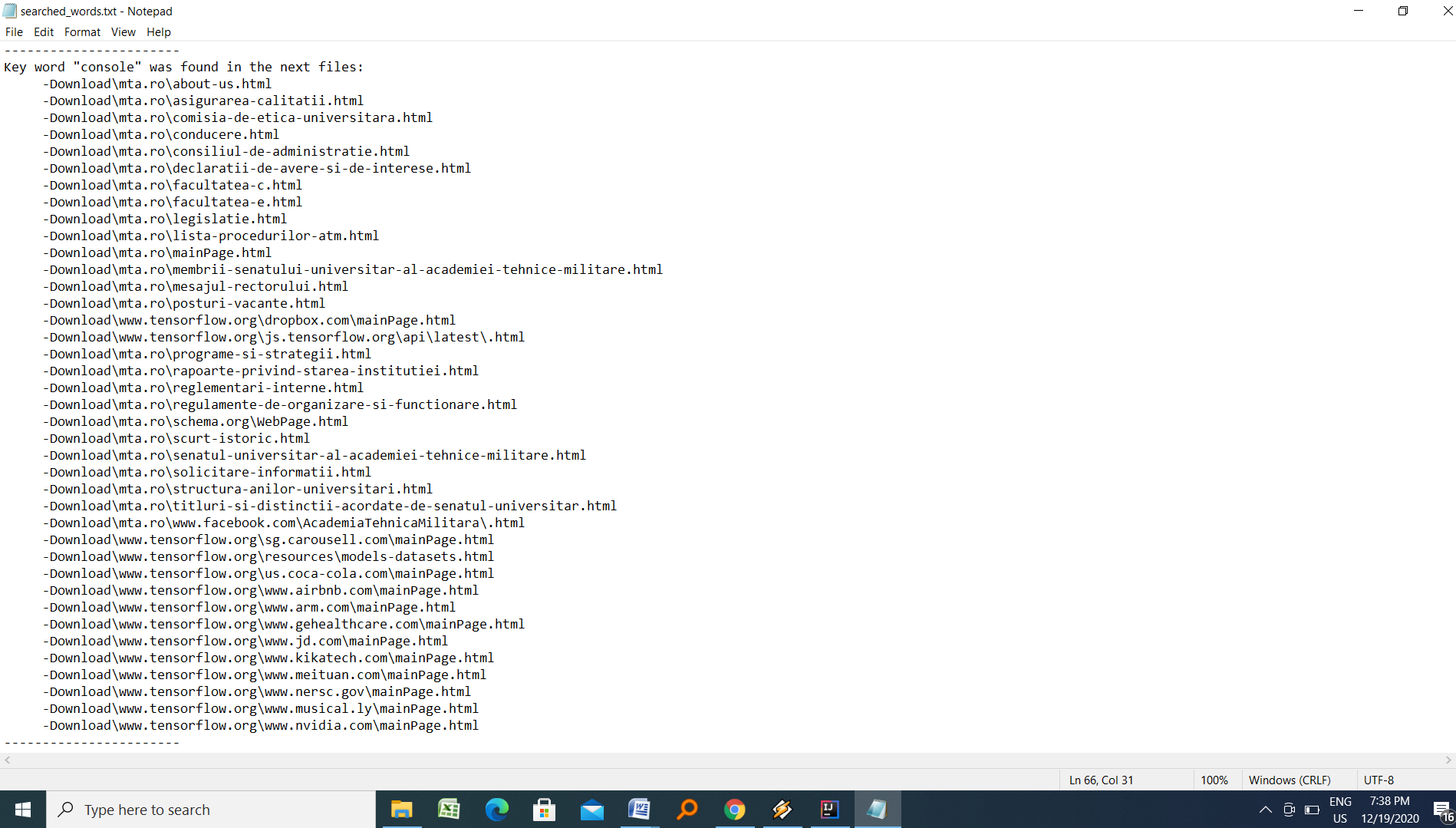


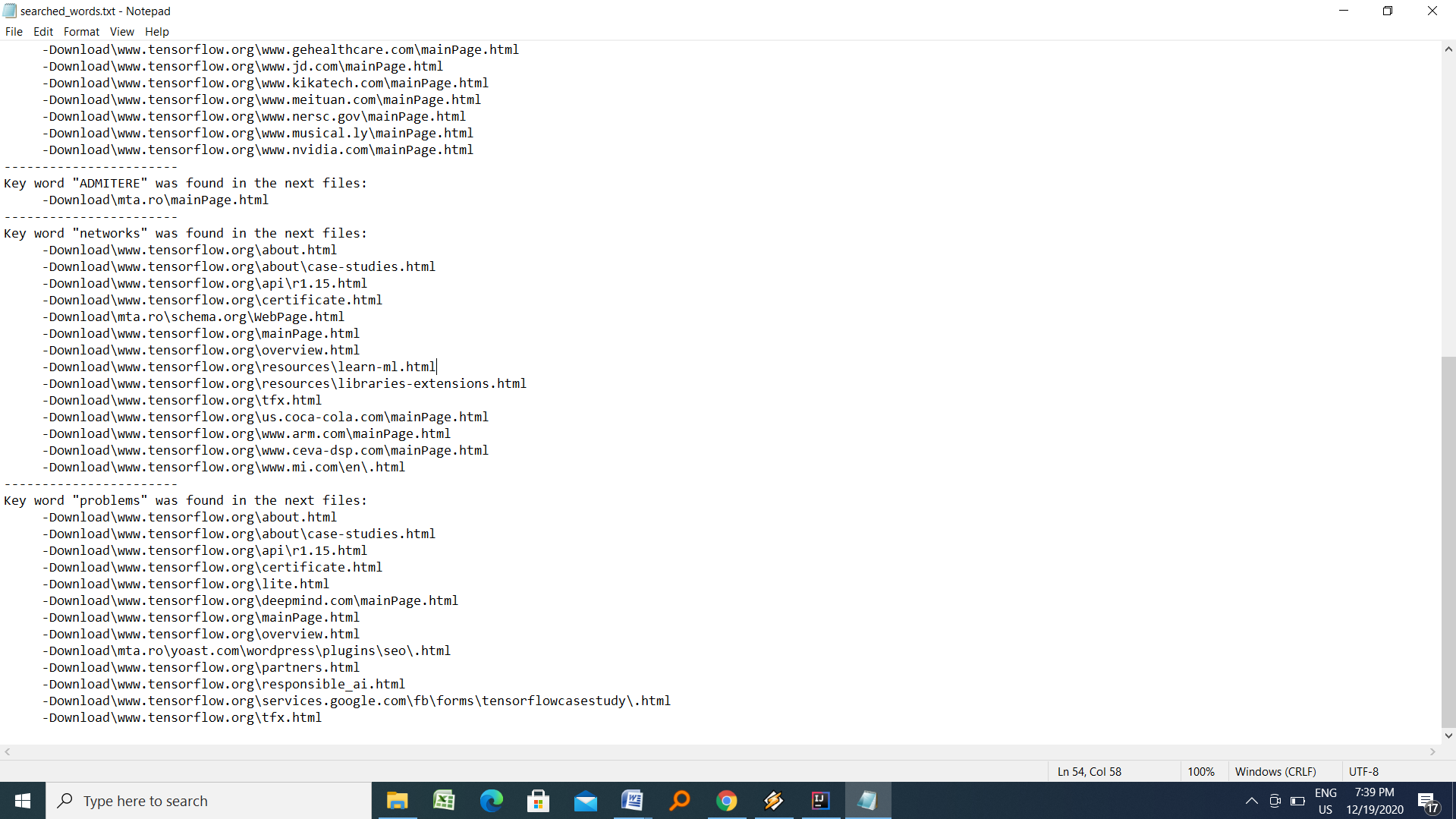


1. Input:

Cuvinte: ADMITERE, STUDENȚI, console, problems, networks

Output(fișierul searched\_words.txt):





**T5. Filtrarea pe baza tipului fișierului**

După ce avem descărcate toate resursele în folder-ul Download, utilizatorul va va avea posibilitatea de a filtra aceste fișiere în funcție de tipul pe care acesta dorește să le păstreze.

Input: se vor păsra doarfișierele cu extensia .txt și .html

Output: Filtrarea s-a efectuat cu succes, doar fișierele tip „.txt” au rămas

Input: se dorește păstrarea fișierelor cu extensia .doc, .docx și .pdf

Output: Filtrarea s-a efectuat cu succes

Input: se dorește păstrarea fișierelor cu extensia .html și .css

Output: Toate fișierele cu extensia menționată au rămas.