

Documentul de specificare a cerințelor software

26 Noiembrie, 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**

Cuprins

1.	Scopul documentului	3
2.	Continutul documentului	3
3.	Descrierea generala a produsului	4
3.1	Situatia curenta	4
3.2	Misiunea proiectului	4
3.3	Contextul proiectului	4
4.	Cerințe funcționale.....	5
4.1	Actori.....	5
4.2	Diagrama de sistem.....	5
4.3	Descrierea cazurilor de utilizare.....	6
4.3.1	Filtrarea datelor în funcție de tip.....	7
4.3.2	Căutarea după cuvinte cheie	7
4.3.3	Limitarea dimenisunii.....	7
4.3.4	Crearea unui sitemap	7
5.	Cerințe nefuncționale	7
5.1	Cerințe de interfață	7
5.2	Cerințe de performanță	7
5.3	Cerințe de securitate	8
5.4	Cerințe de disponibilitate.....	8

1. Scopul documentului

Documentul de specificare a cerințelor software prezent furnizează o descriere completă a unui Web Crawler care furnizează servicii specifice de bază.

În acest document sunt prezentate comportarea externă a aplicației, sunt descrise cerințele funcționale și nefuncționale, restricțiile de proiectare și alți factori necesari pentru descrierea completă a cerințelor software.

2. Continutul documentului

Documentul de față conține cinci capitole, ce descriu progresiv toate cerințele de implementare enunțate.

Primul capitol este o scurtă prezentare a obiectivelor pe care acest proiect și le propune să le realizeze.

Capitolul al doilea reprezintă o scurtă trecere în revistă a informațiilor utile ce se regăsesc în acest document.

Capitolul al treilea își propune să facă o descriere generală a contextului în care a luat ființă acest proiect, cuprinzând situația actuală ce a dus la luarea deciziei de realizare a acestui Web Crawler, precum și scopul acestui proiect prin prisma îndeplinirii cerințelor de implementare exprimate de solicitant.

Cel de-al patrulea capitol prezintă cerințele funcționale ale proiectului. Este ilustrat modul în care utilizatorul interacționează cu sistemul propus de echipa noastră, precum și modul în care decurge procesul de funcționare. Pentru o mai bună înțelegere a modului de interacționare cu soluția propusă de echipa

noastră, este furnizată în acest capitol și diagrama de sistem, ce ilustrează sugestiv comportamentul sistemului.

Ultimul capitol, al cincilea, enumeră cerințele nefuncționale ale soluției propuse. Sunt definite aici necesitățile minimale ale acesui proiect pentru o funcționare optimă, precum și pentru asigurarea integrității datelor transmise și o bună desfășurare a procesului de funcționare.

3. Descrierea generala a produsului

3.1 Situatia curenta

Analizând necesitatea studenților de a avea la dispoziție mai multe tool-uri software pentru eficientizarea activităților din cadrul facultății de calculatoare, s-a constatat faptul că o aplicație de tip Web Crawler ar fi mai mult decât utilă pentru a descărca și indexa conținutul de pe întreg Internetul, în scopul organizării datelor extrase.

3.2 Misiunea proiectului

Proiectul propus are ca scop plasarea studenților în situația practică de dezvoltare a unui proiect software conform standardelor din industrie. Acesta urmărește familiarizarea studenților cu lucrul în echipe de dezvoltare, utilizarea soluțiilor existente în industrie, capacitatea de învățare și de asumare de responsabilități corespunzătoare rolului ocupat în cadrul dezvoltării proiectului.

3.3 Contextul proiectului

Produsul software este conceput pentru furnizarea unor funcționalități de bază ale unui Web Crawler.

Pentru o organizare cât mai eficientă în cadrul proiectului, acesta este împărțit și distribuit în cadrul echipelor de studenți, fiecare având rolul său bine stabilit

în îndeplinirea obiectivului. Un ultim pas în dezvoltarea proiectului îl presupune îmbinarea modulelor pentru obținerea produsului final și funcțional.

Un crawler Web este un bot Internet care ajută la indexarea Web. Acestea accesează cu crawlere câte o pagină pe rând printr-un site web până când toate paginile au fost indexate. Crawlerele web ajută la colectarea informațiilor despre un site web și a linkurilor aferente acestora și, de asemenea, ajută la validarea codului HTML și a hyperlinkurilor.

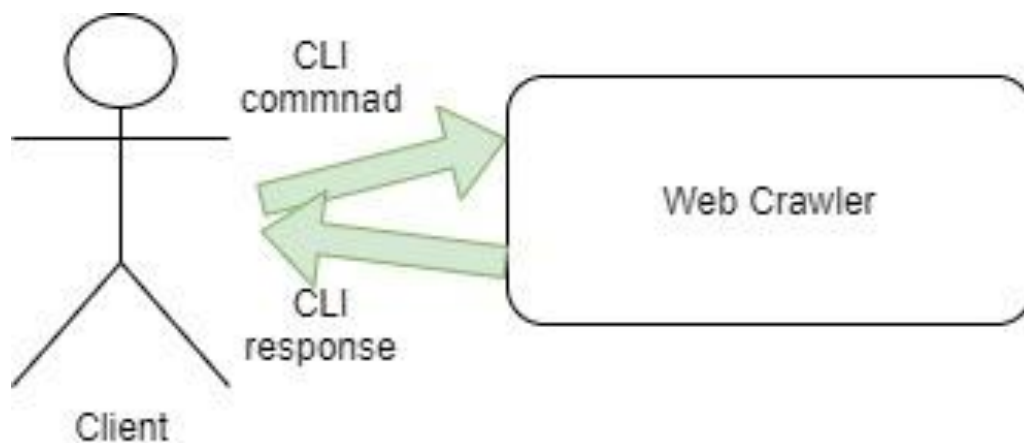
4. Cerințe funcționale

4.1 Actori

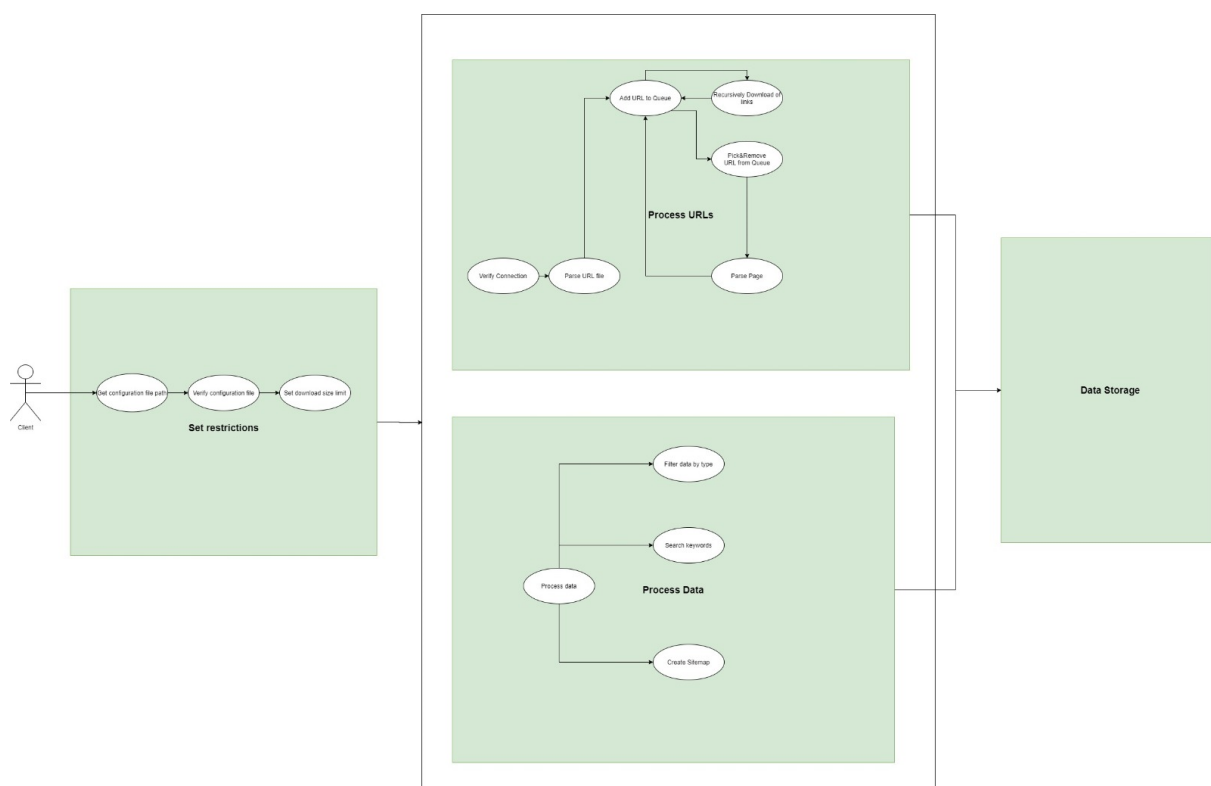
Actorii acestui sistem sunt utilizatorii, care prin intermediul unei linii de comandă pot da comenzi pentru descărcarea resurselor unui site din cadrul unui server web, date care mai apoi pot fi manipulate în funcție de cerințele utilizatorului.

4.2 Diagrama de sistem

Interacțiunea utilizatorului cu produsul software



Interacțiunea între componentele produsului software



4.3 Descrierea cazurilor de utilizare

Inițial, programul va primi un fișier de configurare din partea utilizatorului, prin intermediul căruia clientul va preciza numărul de thread-uri, delay-ul folosit, calea către directorul ce urmează să salveze paginile descărcate și adâncimea până la care se va merge recursiv.

Odată verificat fișierul de configurare, clientul va introduce un calea către un fișier ce conține URL-urile ce se vor descărcate. Se vor citi URL-urile și se vor dispune într-o coadă de așteptare. Apoi, luând pe rând fiecare

URL din coadă, se vor descarca local resursele din pagina respectivă, urmând ca mai apoi să se caute toate link-urile către alte pagini, introducându-le și pe acestea în coadă. Toată această operațiune va fi împărțită pe thread-uri, în funcție de o împărțire egală, evitând astfel o suprasolicitare a unuia din thread-uri.

La finalizarea descărcării resurselor necesare, utilizatorul va fi informat cu privire la rata de success a descărcării, urmând ca mai apoi el să aibă la dispoziție numeroase metode de căutare și analiza a datelor extrase:

4.3.1 Filtrarea datelor în funcție de tip

Utilizatorul este capabil să extragă din fișierele salvate local datele de interes personal. Acesta poate filtra informațiile în funcție de tipul, dimensiunea și data descărcării acestor fișiere.

4.3.2 Căutarea după cuvinte cheie

Pentru un mai bun management al datelor minate, utilizatorul poate căute prin resurse în funcție de niște cuvinte cheie. Astfel, aplicație se va comporta precum un “search engine”, dar numai ca la nivel local.

4.3.3 Limitarea dimensiunii

Utilizatorul are posibilitatea de a seta dimensiunea fișierelor pe care dorește să le descarce, evitând astfel limitarea spațiului disponibil pe disk.

4.3.4 Crearea unui sitemap

Aplicația pune la dispoziție oportunitatea realizării unui sitemap, care constituie un fișier de output ce va conține elementele paginii după tip, dată și link-urile aferente organizate sub forma unei structuri.

5. Cerințe nefuncționale

5.1 Cerințe de interfață

Soluția propusă poate funcționa pe următoarele sisteme de operare: Windows 10, Windows 8.

5.2 Cerințe de performanță

Soluția nu necesită specificații diferite față de cele standard ale sistemului de operare pe care poate funcționa.

5.3 Cerințe de securitate

Sistemul nu necesită măsuri de securitate, altele în afară de cele recomandate pentru sistemul dumneavoastră de operare.

5.4 Cerințe de disponibilitate

Sistemul funcționează atâta timp cât dispozitivul este conectat la o sursă de Internet viabilă.