

Crawling

Fisa cerintelor

1. Descriere
2. Actorii
3. Actionari si interese
4. Scenarii de utilizare
 - a. Extinderea arealului de cautare si extragere
 - b. Depistarea crawler-ului de catre aplicatie
 - c. Intreruperea conexiunii cu aplicatia web
 - d. Intreruperea conexiunii cu baza noastra de date

Echipa

*Andone Madalin
Buruiana Sebastian
Marculet Dan
Nechita Mihai*

Descriere

Modulul de crawling ofera modalitati si functionalitati in ceea ce priveste extractia de date din cadrul unor aplicatii web deja existente din domeniul. Se furnizeaza, asadar, materia bruta pentru modulul de baze de date in scopul obtinerii unui esantion de testare a metodelor de predictie.

Crawler-ul va fi capabil sa abstractizeze datele preluate din cadrul diferitelor site-uri si sa le adapteze la o structura uniforma, omogena. Ulterior, baza de date va fi populata cu elementele relevante din cadrul acestor date.

Actorii

- Aplicatia target
- Web-ul
- Crawler-ul
- Modulul de baze de date (clientul)

Actionari si interese

Crawler-ul va actiona in vederea satisfacerii nevoilor bazei de date prin obtinerea informatiilor de populare necesare. De asemenea, va realiza acest lucru intr-un mod controlat, astfel incat datele sa nu fie aberante. Se va pastra astfel un context in ceea ce priveste cautarea si furnizarea de informatii.

Scenarii de utilizare

Extinderea arealului de cautare si extragere

Obiectiv/Context: Modulul pentru baza de date cauta diversitate in cadrul datelor (e.g se doresc imobile atat din orasul Iasi, cat si din orasul Bucuresti).

Scenariu: Se analizeaza alte surse din cadrul Web-ului in vederea antrenarii Crawler-ului si obtinerea informatiilor necesare

Depistarea crawler-ului de catre aplicatie

Obiectiv/Context: Aplicatia identifica Crawler-ul si il blocheaza

Scenariu: Se utilizeaza un proxy si se cauta intr-un mod neintrusiv datele (prin intermediul unor sleep-uri)

Intreruperea conexiunii cu aplicatia web

Obiectiv/Context: Conexiunea cu aplicatia exploatata de crawler se inchide neasteptat

Scenariu: Crawler-ul reuseste sa salveze datele deja obtinute care nu sunt corupte si incearca reconectarea

Intreruperea conexiunii cu baza noastra de date

Obiectiv/Context: Conexiunea cu modulul de baza de date se intrerupe

Scenariu: Se urmareste reobtinerea conexiunii intr-un timp cat mai scurt si remedierea eventualelor daune