
**UNIVERSITATEA SAPIENTIA DIN CLUJ-NAPOCA
FACULTATEA DE ȘTIINȚE TEHNICE ȘI UMANISTE,
TÎRGU-MUREȘ
PROGRAMUL DE STUDII ...**

Price Monitor

PROIECT DE DIPLOMĂ

Coordonator științific:
Dr. Szántó Zoltán

Absolvent:
Palfi Szabolcs

2021

UNIVERSITATEA “SAPIENTIA” din CLUJ-NAPOCA Facultatea de Științe Tehnice și Umaniste din Târgu Mureș Specializarea: ...		Viza facultății:
LUCRARE DE DIPLOMĂ		
Coordonator științific:	Candidat: Anul absolvirii:	
a) Tema lucrării de licență:		
b) Problemele principale tratate:		
c) Desene obligatorii:		
d) Softuri obligatorii:		
e) Bibliografia recomandată:		
f) Termene obligatorii de consultații: săptămânal		
g) Locul și durata practicii: Universitatea Sapientia, Facultatea de Științe Tehnice și Umaniste din Târgu Mureș		
Primit tema la data de:		
Termen de predare:		
Semnătura Director Departament	Semnătura coordonatorului	
Semnătura responsabilului programului de studiu	Semnătura candidatului	

Declarație

Subsemnatul/a ... , absolvent al specializării ..., promoția ... cunoscând prevederile Legii Educației Naționale 1/2011 și a Codului de etică și deontologie profesională a Universității Sapienția cu privire la furt intelectual declar pe propria răspundere că prezenta lucrare de licență/proiect de diplomă/disertație se bazează pe activitatea personală, cercetarea/proiectarea este efectuată de mine, informațiile și datele preluate din literatura de specialitate sunt citate în mod corespunzător.

Târgu Mureș,

Data:

Extras

Extract

Cuvinte cheie:

**SAPIENTIA ERDÉLYI MAGYAR
TUDOMÁNYEGYETEM
MAROSVÁSÁRHELYI KAR
SZÁMÍTÁSTECHNIKA SZAK**

Price Monitor

DIPLOMADOLGOZAT

**Témavezető:
Dr. Szántó Zoltán**

**Végzős hallgató:
Palfi Szabolcs**

2021

Kivonat

Kivonat

Kulcsszavak:

Abstract

Abstract

Keywords:

Tartalomjegyzék

1. Bevezető	1
2. Célkitűzések	3
3. Szakirodalom áttekintése	5
3.1. Web bányászat	5
3.2. Web Struktúra Bányászat	7
3.2.1. Szemelyre szabott információ lekérése példával illusztrálva	7
3.3. A Web Scraping gyakori alkalmazásai	8
3.4. Web Scraping Technikák	9
3.5. Web Scraping Szoftverek	9
3.6. Legális és Etikai keretek	11
3.6.1. Felhasználói feltételek	11
3.6.2. Szerzői jogok	11
3.6.3. GDPR	12
3.6.4. Weboldal Károsítása	12
4. Rendszer specifikációi	14
4.1. Felhasználói Követelmények	14
4.2. Rendszer Követelmények	17
4.2.1. Funkcionális követelmények	17
4.2.2. Nem-Funkcionális követelmények	18

5. Összefoglalás	20
5.1. Összefoglalás	20
Irodalomjegyzék	20
A. Függelék	22

Ábrák jegyzéke

3.1. A Web Bányászat rendszertana [1]	6
4.1. Az ár változását ábrázoló diagram	15
4.2. Use Case diagram	16
4.3. Use Case diagram	16

1. fejezet

Bevezető

A mai rohanó világban a bevásárlások egyre növekvő százaléka történik az Interneten, mindez lehetőséget nyújtva a vásárlóknak, hogy egy bizonyos terméket több, akár hazai akár külföldi, oldalról is megvásárolhasson. Az e-commerce-el foglalkozó cégek rohamos fejlődésnek indultak az utóbbi évtizedben mely maga után vonja az érdekesebbnél érdekesebb marketing fogásokat, melyekkel a célközönséget próbálják vásárlásra bírni.

Valószínűleg mindenki hallott már a “Black Friday” az-az „Fekete Péntek” -nek nevezett jelenségről amely inspirációként szolgált az alkalmazás megvalósításához. Ez a kifejezés legelőször az 1800-as években fogalmazódott meg, amikor is Jay Gould és James Fisk az amerikai arany árak manipulálása által 20%-os esést okoztak a részvényt piacon melynek következtében az árucikkek értéke felére csökkent ¹. A 20. század közepe fele ez már egészen más jelentéssel bírt, ugyanis a Hálaadás ünnepét követő napon, az-az pénteken vette kezdetét a karácsonyi árleszállítás, mely sok cég esetében életmentő volt, hiszen ekkor kerültek át a veszteséges állapotból melyet pirossal jelöltek, a jövedelmezőbe, amit már fekete írószerrel jegyeztek fel ². Ebben az időszakban a megszokottnál jóval nagyobb és több árleszállítással vonzották az embereket.

Mint azt sokan tudjuk, országunkban is nagy népszerűségnek örvend ez a jelenség, habár eléggé távol áll az eredeti koncepciótól. Nagyon sok mesterséges árleszállítással próbálják becsapni az em-

¹[https://en.wikipedia.org/wiki/Black_Friday_\(1869\)](https://en.wikipedia.org/wiki/Black_Friday_(1869))

²[https://en.wikipedia.org/wiki/Black_Friday_\(shopping\)](https://en.wikipedia.org/wiki/Black_Friday_(shopping))

bert, melyet legtöbb esetben jól kitervelt ár ingadozással oldanak meg³. Ugyanakkor, nem kizárólag ebben a periódusban lehet észrevenni az úgymond „hamis” kedvezményeket ezért szükségét láttuk egy olyan alkalmazás kifejlesztésének, amely nyomon tudja követni egy megadott termék árat, illetve annak ingadozását.

Mivel az Interneten publikus adatok találhatók, ezek felhasználásával semmiéle kár nem keletkezik az adott weboldalak számára. Egy internetes oldal betöltése során, mi, mint felhasználók, egy kérést intézünk egy szerver fele a böngészőnkön keresztül, ami majd a kapott válasz alapján felépíti és megjeleníti számunkra a megtekinteni kívánt oldalt. Ezt a műveletet természetesen legtöbbször ahogy az előbbiekben is említettem, böngészőn keresztül végezzük, viszont ez nem egy szükséglet, inkább egy eszköz, számos más módon is intézhetünk kéréseket egy adott szerver fele. Az általunk megvalósítani kívánt alkalmazás ezt a tulajdonságot hivatott kihasználni, ezáltal nyilvánosan elérhető adatok begyűjtésével, feldolgozásával és elemzésével szeretne foglalkozni.

³<https://cavaleria.ro/tepele-de-black-friday-2020/>

2. fejezet

Célkitűzések

Az alábbi fejezetben összefoglaljuk az alkalmazás fontosabb célkitűzéseit. A fő cél, mely érdekében létrejön a szoftver, az, hogy segítsen egy potenciális vásárlónak követni egy adott termék árának időbeli változását. Ennek megfelelően, a következő célok fogalmazódtak meg:

- Rövid használati útmutató, szöveges formában
- Támogatott oldalak listázása
- Regisztrálási, bejelentkezési lehetőség biztosítása, hogy különböző eszközökön is, mint például mobilos vagy webes, elérhetőek legyenek a követett termékek információi
- Felhasználói fiók jelszavának változtatási lehetősége, felhasználói fiókból való ki jelentkezés valamint annak törlése.
- A termékek ábrázolása listaszerűen történjen
- Az árak változását a felhasználó számára grafikus formában szeretnénk ábrázolni a könnyebb átláthatóság érdekében
- Az árak időben való változását könnyen átlátható vonal diagramon ábrázolni
- Egyszerű átirányítás a termék oldalára
- Telefonos alkalmazás

- Kiegészítő, Chrome alapú böngészőkre

3. fejezet

Szakirodalom áttekintése

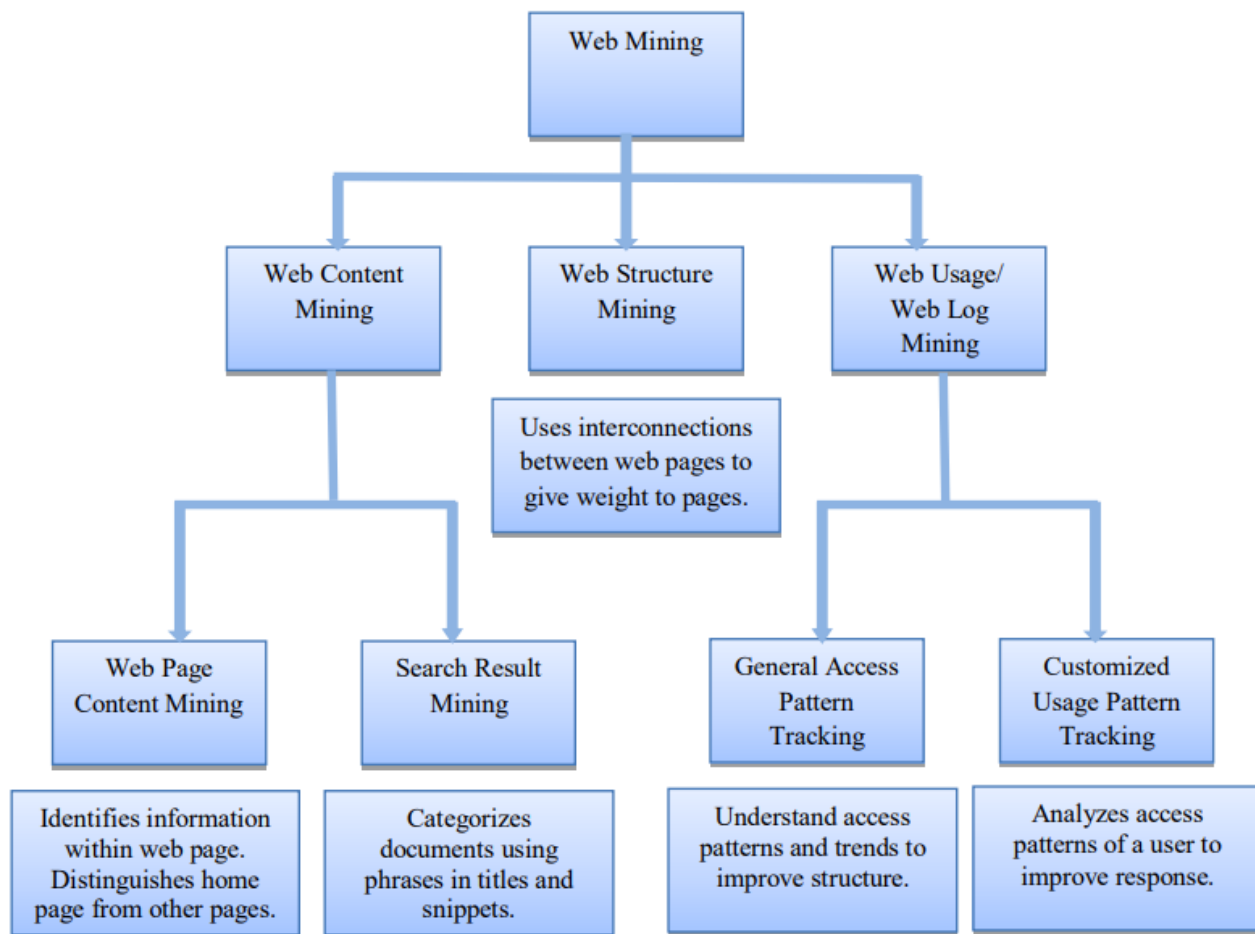
3.1. Web bányászat

Web bányászat, vagy angol kifejezésben Web mining, alatt azt a folyamatot értjük, amely által új, eddig nem ismert, de hasznos információt fedezünk fel az Interneten található adatok között. Mindezt a cégek arra használják, hogy új értékes információkat gyűjtsenek, ezeket feldolgozzák, majd ezek által a felhasználókat vagy fogyasztókat jobban megismerjék. Ez a folyamat az úgynevezett adat bányászat technikát alkalmazza ahhoz, hogy automatikusan kinyerje az adatokat az Internetről[2].

Számos más technikát is alkalmaztak már új információk kinyerésére, az így is hatalmas és egyre növekvő adat mennyiségből, mint például az Information retrieval, Information extraction illetve gépi tanulás. Az Information retrieval működési elve az, hogy a szöveg indexelése után nyeri ki a hasznos információt. Az Information Extraction arra fókuszál, hogy csak a lényeges információt nyerje ki, míg az előbb említett inkább hasznos dokumentumokat jelöl meg. A gépi tanulós módszer nem kötődik direkt módon a Web Scraping-el viszont segítséget nyújt a szövegek osztályozási folyamatában. A web bányászatot három fő kategóriába soroljuk, ahogy az látható az 3.1 ábrán is.

A Web Content Mining vagy magyarul webes tartalom bányászat, olyan tartalmakra fekteti a hangsúlyt, mint például szöveg, kép, audio, video, metaadatok, hiperlinkek. Ezek tanulmányozása, segít nekünk megérteni a felhasználók, vásárlók viselkedését mely által a weboldalak teljesítményét növelni lehet a későbbiekben, hogy azok jobban, illetve hatékonyabban működjenek.

Mivel a webes tartalom bányászat megvizsgálja úgy a kereséseket, mint azoknak a konkrét tartal-



3.1. ábra. A Web Bányászat rendszertana [1]

mát ezért ezt is tovább lehet osztani két kategóriába, Keresési Eredmény, illetve Weboldal Tartalom bányászat. Nevükből adódóan, ezek kiegészítik egymást, mivel a tartalom bányászat azon oldalakon történik, melyeket korábban megvizsgált és ígéretesnek talált a Keresési eredmény általi elemzés. A Web Structure Mining egy olyan ágazat, mely struktúrák bányászásával foglalkozik, mint például HTML vagy XML címkék, amely által weboldalak közötti kapcsolatot tud felismerni, ezáltal súlyokat rendelve azokhoz. A Web Usage Mining által lehet megérteni a különböző használati mintákat, amelyeket a felhasználók követnek, ezeket főként naplózás, felhasználók profiljai, sütik, könyvjelzők által, de ugyanide tartoznak a különböző egér mozdulatok vagy görgetési adatok is.

3.2. Web Struktúra Bányászat

Rengeteg új adat generálódik az Interneten, napról napra az adat mennyiség exponenciálisan növekszik. Bőségesen állnak rendelkezésünkre szolgáltatások, illetve információk, elektronikus áruházak, elektronikus újságok, közösségi oldalak formájában, hogy csak párat említsünk. Habár ezen adatok fogyasztás céljára lettek szánva, sok időt el lehet tölteni az adatok kinyerésével és elemzésével. Továbbá, a weboldalak adatai HTML, illetve más webre szánt formátumban vannak jelen, amely megnehezíti az automata feldolgozást. Ez lett a mozgató rugója az ezen a téren zajló kutatásnak, mint például a Web Scraping.

A Web Structure Mining, vagy ismertebb nevén Web Scraping, az a folyamat, amely által hasznos információkat nyerünk ki egy weboldal HTML kódjából, amely az Internet fő formázási eszköze [3]. Az egyik metodológia megállapította, a rendezett annotációk, melyek nem mások, mint gépek számára is értelmezhető leíró információk az adott oldalra tekintve, egy külön szemantikus rétegben vannak tárolva, elválasztva a weboldaltól, ezáltal megkönnyítve és felgyorsítva a bányászási folyamatot, mivel először ezeket a fontos metaadatokat dolgozza fel, majd csak ezt követően magát a weboldalt.

A HTML struktúráját tekintve, két fő adat típust lehet bányászni belőle. Az egyik a felhasználó által generált - a másik pedig a metaadat. A felhasználó által generált adat bármi olyan típusú adatra vonatkozik, amelyet a felhasználó hozott létre vagy adott hozzá a weboldalhoz, akár személyesen akár egy közösségi platform hozzátartozásával. A metaadat definíció szerint olyan adat, amely leír egy másik adatot [4], ezek általában minden weboldalon megtalálhatók, fontosabb leíró adatokat tartalmazva, mint például szerző, cím, cikkek esetén megjelenés időpontja, kulcsszavak. Ezek általában nem láthatóak a felhasználó számára, viszont kiolvashatóak az adott oldal HTML kódjából[4].

3.2.1. Szemelyre szabott információ lekérése példával illusztrálva

Tegyük fel, hogy egy személy fárasztónak találja, hogy a nap végen a fontos vagy számára értékes hírek után keressen, vagy előkeresse a kedvenc sport csapata elért eredményeit. Egy intelligens Web Scraper a tökéletes eszköz ebben az esetben, mivel az időközönként végig tudja böngészni az Internetet a felhasználó által megszabott témakörökben található információk után kutatva, vagy akár specifikus kulcsszavakat tartalmazó híreket előkértivé. Ez egy előre weboldalakat tartalmazó listán

menne végig, melyet a felhasználó határoz meg, hogy a számára hiteles információt kapja meg[3].

Egy másik példa egy Web Scraper felhasználására az, amikor például egy vásárló több terméket is kinézett magának, több különböző weboldalon. Ha esetleg nem szeretné azonnal megvásárolni a terméket, hanem csak követni annak árát, akkor esettől függően, több oldalra is be kell jelentkeznie, több felhasználóval oldalanként, de legjobb esetben is számos oldalt kellene naponta megfigyeljen és lejegyezzen. Egy Web Scraper abban könnyítené meg a felhasználó dolgát, hogy például egy böngészős kiegészítő keretein belül, a vásárló hozzá tudja adni egy listához a terméket és attól a pillanattól, a program naponta akár többször is le tudja kérni a termék árát, anélkül, hogy a felhasználó bármit is tenne. Ezáltal egy helyen lenne a vásárló több terméke, és pár kattintással tisztább képét alkothat a termékek árára vonatkozóan

3.3. A Web Scraping gyakori alkalmazásai

- Online ár összehasonlítás – ugyanazon termék árának összehasonlítása több weboldalon, pl. <https://www.compari.ro/>, <https://www.price.ro/>
- Contact Scraping – általában email címeket gyűjtenek, marketing céljából
- Időjárással kapcsolatos adatok gyűjtése
- Weboldal változásainak figyelése
- Több forrásból származó adat egyesítése
- Kedvezmény kuponok pl. pouch – chrome extension
- Álláshirdetések összesítése
- Brand monitoring – egy bizonyos márkához tartozó adatokat gyűjtik, általában az ahhoz társított véleményre kíváncsiak
- Piac tanulmány – Egy adott termék piacon való elhelyezkedését, illetve potenciális sikerességet próbálják megjósolni a bányászott adatok átvizsgálásával.

3.4. Web Scraping Technikák

Számos technika áll rendelkezésünkre melyekkel az adatgyűjtést végezhetjük, ezeket mindig az adott helyzetnek megfelelően kell kiválasztani, főként a hatékonyságot tartva szemelőt. Ebben a részben a Web Scraping egy pár technikája kerül röviden ismertetésre.

Copy-paste - Időközönként valaki kézzel történő adatgyűjtést, valamint vizsgálatot végez. Adott helyzetekben ez a leghatékonyabb módszer, viszont nagyon hajlamos a hibákra, sok időt és fáradságot vesz igénybe az ember részéről, amíg a nagy adathalmazokat feldolgozza.

Reguláris kifejezések - Ez egy egyszerű és erőteljes megközelítése az információ gyűjtésnek. A UNIX vagy más programozási nyelv által használt reguláris kifejezés illesztésen alapszik.

HTML Parsing – Félig strukturált lekérdező nyelvek segítségével elemezni, illetve módosítani a weboldalak tartalmát.

DOM Parsing – A böngészőkbe beépített kezelő programok segítségével, az erre a célra fejlesztett alkalmazások lekérhetik a kliens oldalon dinamikusan létrejött tartalmakat is, mellyel utána fel tudják építeni a DOM fát, ebben pedig könnyebben lehet specifikus adatok után keresni.

Web Scraping Szoftver – Számos szoftver áll rendelkezésünkre, amelyeket személyre szabott keresésre lehet használni.

Mesterséges Intelligencia – Több helyen is kísérleteznek gépi tanulós adatbányászattal, melynek az a célja, hogy a gépek megtanulják úgy értelmezni a weboldalakat, ahogy azt az emberek tennék.

3.5. Web Scraping Szoftverek

A Web Scraping szoftverek rendkívül fontos szerepet játszanak ezen a téren, mivel automatizálják és rendkívül felgyorsítják az adat-gyűjtő, valamint feldolgozó folyamatot. Számos ilyen szoftver található a piacon, a maga előnyeivel és hátrányaival. Ezeknek az ára a funkcionalitásuk függvényében, valamint a támogatás és frissítési időszakok függvényében változik.

Visual Web Ripper¹ – Az egyik legfejlettebb web scraping szoftver, melyet a Sequentum csoport fejlesztett 2006-tól kezdődően. Weboldalokról gyűjtött információk bányászására használják, úgy egy-

¹<http://visualwebripper.com/>

szerű weboldalak, mint e-commerce oldalak esetében, mint például eBay, Amazon, magento, azonban titkosított tartalmak esetében is segítségünkre lehet. A bányászott adatokat kimenthetjük adatbázisba vagy CSV, illetve XML formátumban is. Előnye, hogy vizuális felülettel rendelkezik, ezért rendkívül egyszerűen ki lehet választani, hogy pontosan mit is szeretnénk. Egyszeri fizetéssel lehet megvásárolni a szoftvert, \$349.00 áron², viszont fontos szempont, hogy ezen alkalmazás elveszti a gyártó általi támogatottságát 2021-gyel kezdődően, kivételt képezve azon esetek, ahol a vásárlóval karbantartási szerződés van érvényben, mely túlhaladja ezt az időpontot.

Web Content Extractor³ – Nagyon jó automatizálási lehetőséget nyújt, rendkívül egyszerűen használható, pár kattintással meg lehet adni a kívánt mintát, ami szerint majd adatokat fog gyűjteni. A program rugalmas, abból a szempontból, hogy nem próbálja túlozkoskodni a felhasználót, hanem egy előlnézetet ad az eredményről, majd a felhasználó maga végezheti el a szükséges módosításokat, amennyiben szükség van erre. Ezt a szoftvert, bérlet alapú előfizetéssel lehet megszerezni, több változat is elérhető, az árak \$30 - \$150 / hónap² között mozognak.

Mozenda⁴ – Az egyik legegyszerűbben használható szoftver ezen a téren, ami lehetővé teszi a kevésbé technikailag hozzáértő személyeknek is, az egyszerű bányászásokat. Fő különbsége más alkalmazásokhoz képest, hogy maga az adatbányászati folyamat a felhőben történik és nem a felhasználó erőforrásait felhasználva, amely hatalmas előnyt jelenthet. Elérhető egy 30 napos próba csomag is, ami után \$250/hónap-tól² kezdődően változnak az árak, a választott csomag függvényében.

Screen-Scraper⁵ – Nagyon fejlett web scraping alkalmazás, amelyet több változatban is el lehet érni. Az alapszínűt verzió ingyenes, ezzel egyszerűbb adatok után lehet bányászni, könnyen kezelhető, nem kell sok technikai tudás hozzá. Más változatok, mint például a profi vagy vállalkozás szintű verziók már sokkal komplexebbek, több lehetőséget nyújtanak. Nagy előnye, hogy más rendszerekkel könnyen összehíható pl. Java, ezért fel lehet használni más, nagyobb szintű programokban is.

Természetesen sok más program is rendelkezésünkre áll, melyek hasonló funkcionalitásokkal rendelkeznek, minden esetben az adott alkalmazásnak megfelelő és legjobban illőt érdemes választani a nagyobb hatékonyság érdekében. Egyéb web scraper szoftverek[5]: WebHarvy, Easy Web Extract,

²Az árak aktualitásáért lásd a szolgáltató weboldalát

³<https://www.webcontentextractor.com/>

⁴<https://www.mozenda.com/>

⁵<https://www.screen-scraper.com/>

WebSunDew, FMiner, Scrapy, import io.

3.6. Legális és Etikai keretek

Ebben a fejezetben a legális valamint etikai kérdésekről lesz szó, amely elég megosztó, illetve nem teljesen egyértelmű terület. Egyenesen a Web Scraping-et nem szabályozza semmilyen törvény, de több területen is problémába lehet ütközni, ilyen például a védett tartalom, az úgynevezett szerződésszegés vagy GDPR. Attól függően, hogy a bányászat milyen országhoz tartozó területen zajlik, figyelembe kell venni az ottani törvénykezést is, ezért is nehéz konkrétan jellemezni legalitási szempontból. A legpontosabb jellemzés talán az lenne, hogy van, ami igen és van, ami nem.

3.6.1. Felhasználói feltételek

Ugyanúgy, mint egy szoftver vagy szolgáltatás esetén, amikor egy weboldalt használunk el kell fogadnunk bizonyos felhasználói feltételeket, melyek leggyakrabban kis felrúgó ablakként jelennek meg, amikor először látogatunk a weboldalra, vagy ezeket a regisztrálás folyamán tudatosítják velünk. Amennyiben valamilyen módon megszegjük ezeket a feltételeket, érvénybe lép a fentebb említett szerződésszegés. Mivel ez csak akkor érvényes, ha a felhasználó explicit elfogadja a feltételeket, jogi szempontból nem nehéz kizárni a Web Scrapinget.

3.6.2. Szerzői jogok

Bányászni vagy újra publikálni olyan adatokat vagy információkat, amelyeknek a szerzője explicit módon fenntartja a szerzői jogokat, legális szempontból szerzői jogsértésnek minősül. Ugyanakkor egy weboldal nem minden esetben rendelkezik a felhasználói által generált adatokkal, vegyünk példának egy film értékelő oldalt, ahol a felhasználók kifejtik a véleményüket [6]. Más ebbe a kategóriába tartozó tartalmak például a videók, képek, zenék, adatbázisok, cikkek. Maga a bányászása ezeknek az adatoknak nem teszi illegálissá, a felhasználási módjuk határozza meg azt, hogy milyen kategóriába soroljuk azt.

3.6.3. GDPR

Az Európai Unió által érvénybe léptetett Általános Adatvédelmi Rendelet⁶ alapján nem tiltott az adatbányászat, kivételt képezve, ha ez a tevékenység nem tartalmaz személyes adatokat. Ilyennek minősül a név, lakcím, email cím, telefonszám, bankkártya adatok, banki adatok, IP cím, születési dátum, foglalkoztatási információk, orvosi adatok, személyes fotók vagy videók.

3.6.4. Weboldal Károsítása

Ha bármilyen tevékenység által, amelyet a Web Scraper végez, túlterheljük az adott weboldal szervereit vagy bármilyen módon sértjük, gátoljuk annak működését bűncselekménynek számít és legális következményei lehetnek. Ehhez azonban a kárnak anyagnak kell lennie, valamint könnyen bizonyíthatónak bíróság előtt, ahhoz, hogy bármiféle kártérítési kérelemre legyen jogosult a weboldal[6].

A fentieket figyelembe véve tehát, nem lehet egyértelmű választ adni arra, hogy a Web Scraping legális-e vagy sem. Helyette a válasz az, hogy helyzettől függ. Maga a Web Scraping nem illegális, viszont akár az is lehet a következő három dolog függvényében[7].

- A bányászott adat típusa
- Hogyan lesz felhasználva a bányászott adat?
- Hogyan lett bányászva az adat?

A bányászott adat típusa szerint két formájára kell nagyon odafigyelni, Személyes Adatok és Szerzői Jogokkal rendelkező adatok, ezek fentebb részletesebben is tárgyalva voltak.

Ahhoz, hogy eldöntsük az esetünk legalitását meg kell vizsgálni, hogy az adat, amit szeretnénk bányászni publikusan elérhető-e vagy sem. Ha az adat eléréséhez nem szükséges bejelentkezni egy adott oldalra akkor a felhasználói feltételek nem érvényesek ezért legálisan lehet bányászni, mivel ez az adat publikusnak számít. Ha bejelentkezésre van szükség a bányászni kívánt adat eléréséhez, akkor a felhasználói feltételek tanulmányozásával el kell dönteni, hogy legális-e vagy sem, mivel azok elfogadásával legálisan alkalmazhatóvá vált számunkra, azaz a megszegésük jogi következményeket vonhat maga után.

⁶<https://gdpr-info.eu/>

Leegyszerűsítve, három alapvető kérdésre kell választ adni, ahhoz, hogy eldöntsük legális-e az adat bányászat, amelyet végre szeretnénk hajtani:

- Személyes adatot bányászok?
- Szerzői jogokkal rendelkező adatot bányászok?
- Bejelentkezés által elérhető adatot bányászok?

Amennyiben mindhárom kérdésre nemleges a válasz, legálisnak bizonyul adott esetben a Web Scraping. Azonban, ha bármelyik kérdésre pozitív választ adunk nagy valószínűséggel újra kell gondolni és figyelmesebben megvizsgálni az eset legális mivoltát. Ugyanakkor egyik esetben sem szabad megfeledkezni arról, hogy minden ország másképp kezelheti ezen kérdéseket, ezért ilyen értelemben is vizsgálnunk kell.

4. fejezet

Rendszer specifikációi

A jelen dolgozatban egy árakat figyelő alkalmazás készítéséről van szó, mellyel különböző termékek árát lehet követni, bizonyos webshopokon. Az alkalmazás több platformon is elérhető, hogy minél könnyebben eljusson a kívánt információ a felhasználóig. A következőkben tárgyalva lesznek a felhasználói, valamint rendszer követelmények is.

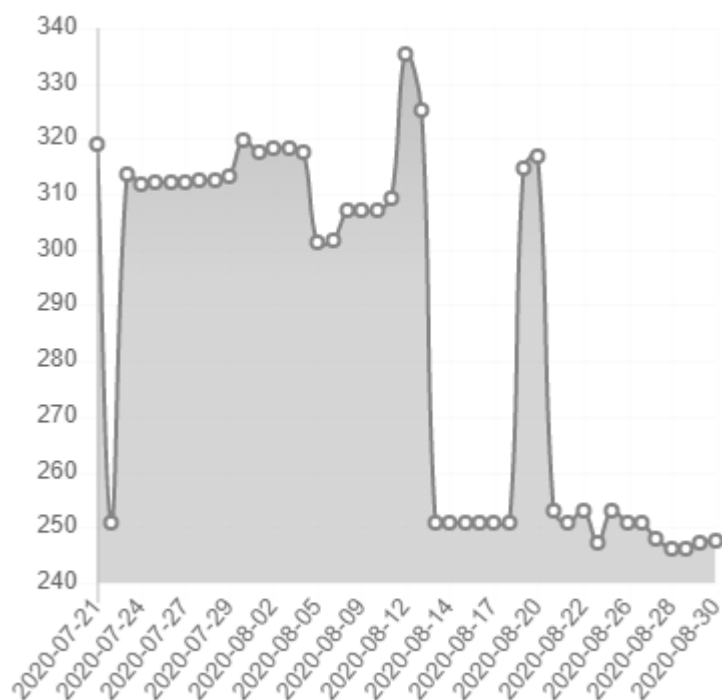
4.1. Felhasználói Követelmények

A felhasználónak mindenekelőtt be kell tudnia jelentkezni, ahhoz, hogy elérje a terméklistáját, ez azért fontos, mivel eszköz váltás esetén nem szeretnénk elveszíteni az addig követett termékeket. A bejelentkezéshez szükséges adatok a regisztrációkor megadott email cím, illetve jelszó. A felhasználó jelszava titkosítva kell legyen tárolás előtt, valamint lehetősége kell legyen igényelni annak változtatását.

Abban az esetben, ha a felhasználó nincs regisztrálva, természetesen ez a funkcionalitás is rendelkezésre áll. A regisztrációhoz szükséges egy érvényes e-mail cím, valamint jelszó megadása. Sikeres regisztrálás esetén a felhasználó vissza kerül a bejelentkező ablakra, ahol be tud lépni, azzal a feltétellel, hogy az automatikusan küldött levél által visszaigazolta email címét. A bejelentkezést követően a felhasználó a főoldalra kerül, ahol a követett termékek listáját tekintheti meg, valamint ezen az oldalon lehetősége van a új termékeket hozzá adni a listához.

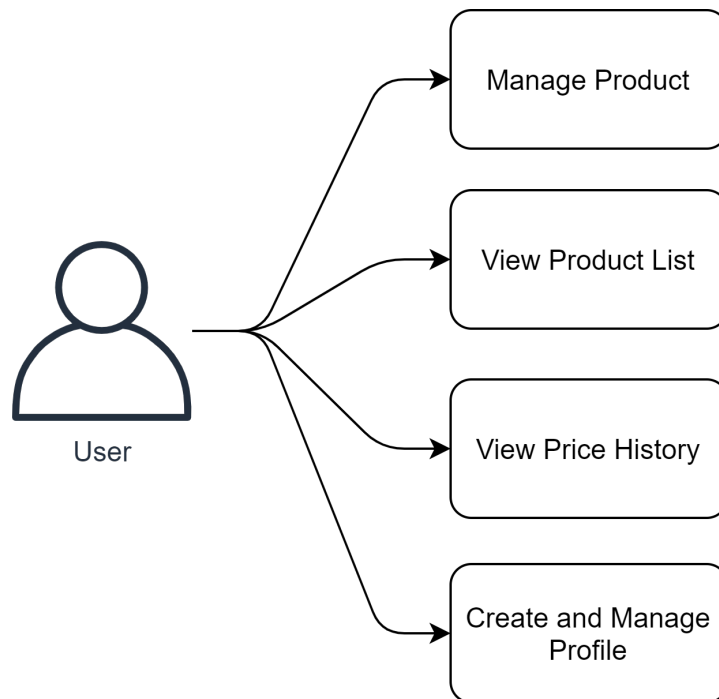
A felhasználó egy grafikonon tekintheti meg az adott termék árának változását a hozzáadás napjá-

tól az aktuális dátumig, lásd 4.1 ábra. Ha a felhasználó úgy dönt, hogy érdekli a termék, lehetősége kell legyen megnyitni az adott oldalt. Amennyiben a felhasználó már nem érdekelt a termék követésében, törölheti az adott terméket, így az többet nem fog szerepelni a listájában.

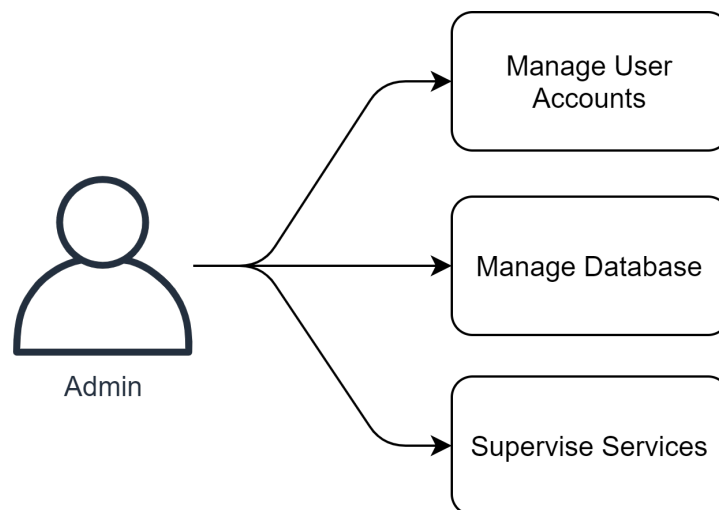


4.1. ábra. Az ár változását ábrázoló diagram

A felhasználónak rendelkezésre kell bocsájtani egy információs részt, amely röviden leírja az alkalmazás használatát, illetve tartalmazza az általa támogatott weboldalak listáját. Továbbá, hasonló módon, a fiókjával kapcsolatos információkat és funkciókat is el kell érnie. Itt tekintheti meg a felhasználó, hogy milyen email címmel jelentkezett be, törölheti a fiókját, illetve kijelentkezhet az alkalmazásból.



4.2. ábra. Use Case diagram



4.3. ábra. Use Case diagram

4.2. Rendszer Követelmények

4.2.1. Funkcionális követelmények

A rendszernek mindenképp előtte, egy bejelentkezési, illetve, regisztrációs felülettel kell rendelkeznie. Regisztrálás után, a felhasználónak egy ellenőrző email-t kell kapnia, amivel igazolja, hogy ő a cím tulajdonosa. A bejelentkezés nem lehetséges, abban az esetben, ha a felhasználó nem igazolta vissza az előbb említett email-ben a címét. A cím igazolása egy linkre való kattintással történik.

A felhasználónak lehetősége van a jelszavának módosítására melyet a bejelentkezési felületről ér el. Miután a felhasználó beírta az email címét, egy levelet fog kapni az adott címre, amelyen keresztül új jelszót tud beállítani.

Bejelentkezést követően, a felhasználó egy felületet lát, melyen bizonyos műveleteket végezhet. Megtekintheti a profiljához tartozó email címét, valamint törölheti a felhasználóját. Ugyanakkor, lehetősége van kijelentkezni az alkalmazásából melynek hatására újra a bejelentkezési oldalra kerül.

Ugyancsak a főoldalról a felhasználónak lehetősége van az alkalmazás használatával kapcsolatos információk megtekintésére mely tartalmaz egy listát is. A lista bizonyos weboldalkát tartalmaz, melyeket kiválasztva, az alkalmazás átirányít az adott elem oldalára.

A felhasználónak lehetősége van termékeket hozzáadni és kitörölni a listájából, valamint görgetni a lista tartalmában. Amennyiben frissíteni szeretné a lista tartalmát, ez úgy lehetséges, hogy a lista tetején tartózkodva, annak tartalmát megpróbálja lehúzni. A terméklistában egy elemet kiválasztva, részletes reprezentációt kap az adott elem tárolt adatairól.

Amennyiben egy új termék került hozzáadásra vagy a termék törlése következett be, a rendszernek ezt fel kell ismernie és elvégeznie a szükséges lépéseket annak érdekében, hogy a felhasználó számára mindig a legfrissebb adatok legyenek láthatóak.

A rendszer back-end részének autonóm módon kell működni, minimális beavatkozással. Amikor egy felhasználó hozzá szeretne adni egy terméket, a rendszer azonnal reagáljon erre a kérésre. A termékek időszakos ellenőrzése is autonóm módon történjen, a megadott időszakokban.

4.2.2. Nem-Funkcionális követelmények

A rendszer backend része, mely a felhasználók által hozzáadott termékek árainak ellenőrzését, frissítését, hozzáadását végzi, megszakítás nélkül kell működnie, napszaktól függetlenül. A rendszernek képesnek kell lennie Python 3-as kódot futtatnia, fel kell legyen telepítve a Python 3.8.0. Létfontosságú, hogy egy stabil Internetkapcsolattal rendelkezzen, vagyis a kapcsolatban ne legyenek megszakítások, valamint a fel-letöltési sebesség ne csökkenjen 10 Mb/s alá, annak érdekében, hogy megfelelő sebességgel lehessen az adatfeldolgozást, valamint az adatok adatbázisba való fel-, letölteset elvégezni. Amennyiben a kapcsolat megszakad, a rendszer azonnal próbáljon újrakapcsolódni, mindaddig amíg ez a művelet nem sikers.

A rendszer vizuális felülete két alapvető platformon kell elérhető legyen. Az egyik egy böngésző kiegészítő, mely bármilyen Chromium alapú böngészőre telepíthető, asztali, valamint hordozható számítógépek esetében is, legalább 58-as verziójú Chrome-ot támogatva. Természetesen ebben az esetben is elengedhetetlen az Internethez való csatlakozás.

A másik egy telefonos alkalmazás, mely Android operációs rendszerrel felszerelt készülékeken legyen elérhető. Az alkalmazásnak legalább 7.0 verziójú Androiddal felszerelt telefonokon kell működnie, mely rendelkezik stabil Internetkapcsolattal. Rendelkeznie kell Light illetve Dark móddal is, melyek közötti váltást automatikusan végzi, az operációs rendszer beállításaihoz igazodva. A felhasználó számára, a követett termékeit egy görgethető listában kell megjeleníteni. Egy listaelem tartalmazza a termék megnevezését, aktuális árát, illetve egy képet róla. A termék ára mellett egy nyíl található, mely azt jelzi a felhasználó számára, hogy az aktuális ár hogyan változott a korábbi ellenőrzéshez képest. A termék árának csökkenését, egy lefele irányuló, a növekedését egy felfele irányuló, amennyiben pedig az nem változott, egy vízszintesen irányított nyíl jelzi. Ezeket a változásokat színekkel is kell jelezni, mely az árra és az előbb említett nyílra hat ki. Csökkenés esetén zöld, növekvés esetén piros, valamint, ha változatlan, akkor fehér vagy fekete színnel kell jelölni ezeket.

A rendszer a Firebase által biztosított Realtime Database nevű adatbázist kell használnia az adatok tárolására. A felhasználó bejelentkezését és a fiókjához tartozó műveleteket szinten a Firebase által biztosított Authentication szolgáltatás végezze, mivel ez biztonságos módon tárolja a szükséges információkat, ugyanakkor, a jelszavakat titkosítva kezeli. Továbbá, ezen keresztül lehessen új jelszót

beállítani, a felhasználót törölni, ugyanakkor a regisztrálás után szükséges email visszaigazolása is ezen a szolgáltatáson keresztül történjen, mivel ezekre egyszerűen használható, ugyanakkor hatékony és biztonságos megoldást nyújt.

Az alkalmazás könnyen használható kell legyen, a felhasználónak minden funkciót három kattintáson belül el kell érnie.

Amikor a felhasználó egy új terméket szeretne hozzáadni a listájához, ne keljen több mint 5 másodpercet várnia ahhoz, hogy az új termék megjelenjen a listájában.

5. fejezet

Összefoglalás

5.1. Összefoglalás

Irodalomjegyzék

- [1] F. Johnson and S. K. Gupta, „Web content mining techniques: a survey,” *International Journal of Computer Applications*, vol. 47, no. 11, 2012.
- [2] F. Gorunescu, *Data Mining: Concepts, models and techniques*, vol. 12. Springer Science & Business Media, 2011.
- [3] B. G. Dastidar, D. Banerjee, and S. Sengupta, „An intelligent survey of personalized information retrieval using web scraper,” *International Journal of Education and Management Engineering*, vol. 6, no. 5, pp. 24–31, 2016.
- [4] R. N. Landers, R. C. Brusso, K. J. Cavanaugh, and A. B. Collmus, „A primer on theory-driven web scraping: Automatic extraction of big data from the internet for use in psychological research,” *Psychological methods*, vol. 21, no. 4, p. 475, 2016.
- [5] D. S. Sirisuriya *et al.*, „A comparative study on web scraping,” 2015.
- [6] V. Krotov and L. Silva, „Legality and ethics of web scraping,” 2018.
- [7] D. Ni, „Is web scraping legal?.”

A. függelék

Függelék