

# Analysis Of Automatic Web Data Collection (Web Scraping) Technologies

**Sharipova Shahodat Bahodir qizi**

Teacher: DSc associate professor **Qodirov Farrukh Ergash ogli**

Master's student of Shahrizabz State Pedagogical Institute

[shahodatsharipova04@gmail.com](mailto:shahodatsharipova04@gmail.com)

## Annotation

Automatic web data collection (web scraping) is one of the important technologies for obtaining and analyzing information in today's digital age. Billions of data are stored on the Internet, but collecting them manually is time-consuming and error-prone. Web scraping technologies allow you to automatically extract data from websites. This article examines the basic definition, technical foundations, areas of application, advantages and disadvantages of web scraping. Libraries widely used for web scraping in programming languages such as Python, JavaScript, Java were analyzed. The article also addresses important questions about web scraping, including legal and ethical issues, robots.txt, API implementation, and data security. The results of the study show that, when used correctly and transparently, web scraping technologies can be a very effective tool for web applications, data analysis, sample research, and market monitoring.

**Keywords:** web scraping, web data collection, automation, data extraction, Python, BeautifulSoup, Selenium, web parsing, data security, digital technologies

## Annotatsiya

Veb-ma'lumotlarni avtomatik yig'ish (web scraping) bugungi raqamli davrida ma'lumot olish va tahlil qilishning muhim texnologiyalaridan biri. Internet-tarmoqda milyardlab ma'lumotlar saqlanib turibdi, lekin ularni qo'lda yig'ish vaqt sarfi va xatolardan iborat. Web scraping texnologiyalari avtomatik ravishda veb-saytlardan ma'lumotlarni ekstraktatsiya qilish imkonini beradi. Ushbu maqolada web scraping-ning asosiy ta'rifi, texnik asoslari, qo'llanilish sohalari, afzalliklari va xavflari o'rganiladi. Python, JavaScript, Java kabi dasturlash tillarida web scraping uchun keng qo'llaniladigan kutubxonalar analiz qilindi. Maqolada web scraping-ning huquq va etika masalalari, robots.txt, API qo'llash va ma'lumot xavfsizligi haqida muhim savollari ham o'rin olgan. Tadqiqotning natijalari ko'rsatadiki, to'g'ri va shaffof tarzda qo'llanganda web scraping texnologiyalari veb-aplikatsiyalar, ma'lumot tahlili, tanlama tadqiqotlari va bozor monitoring uchun juda samarali vosita bo'lishi mumkin.

**Kalit so'zlar:** web scraping, veb-ma'lumotlarni yig'ish, avtomatsiyalash, ma'lumot ekstraktatsiyasi, Python, BeautifulSoup, Selenium, veb-parsing, ma'lumot xavfsizligi, raqamli texnologiyalar

## 1. Kirish

Zamonaviy raqamli davrda internetdagi ma'lumotlarning eksponensial o'sishi tadqiqotchilar, korxonalar va ma'lumotlar olimlari uchun ham imkoniyatlar, ham muammolarni keltirib chiqaradi. Har kuni yaratilayotgan ma'lumotlar hajmi misli ko'rilmagan darajaga yetdi; taxminlarga ko'ra, har kuni taxminan 2,5 kvintillion bayt ma'lumot yaratilmoqda. Biroq, bu ulkan raqamli makondan mazmunli ma'lumotlarni ajratib olish murakkab vosita va metodologiyalarni talab etadi. Veb-manbalardan avtomatlashtirilgan ma'lumot yig'ish uchun eng keng tarqalgan texnologiyalardan biri veb-scraping, ya'ni veb ma'lumotlarini ajratib olish yoki veb hosil qilish deb ataladi.

Veb-scraping maxsus dasturiy vositalar yoki dasturlash skriptlari yordamida veb-saytlardan strukturaviy va no-strukturaviy ma'lumotlarni avtomatlashtirilgan tarzda olish jarayonini anglatadi. Veb-saytlardan ma'lumotlarni qo'lda nusxalash va joylashtirish o'rniga, veb-scraping dasturchilar va ma'lumot tahlilchilariga veb-sahifalarga dasturiy ta'minot orqali kirish, HTML yoki XML hujjatlarni tahlil qilish va tegishli ma'lumotlarni strukturaviy formatda olish imkonini

beradi. Ushbu texnologiya elektron tijorat, moliyaviy xizmatlar, ko'chmas mulk, sog'liqni saqlash va akademik tadqiqotlar kabi turli sohalarda tobora muhim ahamiyat kasb etmoqda. Bir nechta veb-manbalardan katta hajmdagi ma'lumotlarni tezda to'plash va qayta ishlash imkoniyati tashkilotlarning bozor tahlili, raqobatbardoshlik bo'yicha razvedka va ma'lumotlarga asoslangan qaror qabul qilish usullarini inqilobiy darajada o'zgartirdi.

Veb-ma'lumotlarni avtomatik yig'ish (web scraping) texnologiyalari bugungi ma'lumot bazasi va raqamli transformatsiya davrida qancha muhim ekanligini ko'rsatish kerak. Veb-texnologiyalarning rivojlanishi bilan birga, Internet-tarmoqda saqlanayotgan ma'lumotlarning hajmi eksponensial ravishda o'sib bormoqda. Har bir sekundda millionlab ma'lumotlar yangi veb-saytlarga qo'shilib, zamonaviy amaliyotlar uchun bu ma'lumotlarni samarali qayta ishlash zaruriyati paydo bo'ldi.

Web scraping - veb-saytlardan avtomatik ravishda ma'lumotlarni ekstraktatsiya qiluvchi texnologiya - bu muammoni yechishning eng samarali usullaridan biri. Python-da BeautifulSoup va Selenium kabi kutubxonalar, JavaScript-da Cheerio va Puppeteer, Java-da Jsoup kabi dasturlash alatlari web scraping-ni sodda va kuchli qildi. Lekin, web scraping texnologiyalarining qo'llanilishida huquq va etika masalalari, ma'lumot xavfsizligi va veb-sayt xizmatining uzilishi haqida jiddiy savollari mavjud.

Ushbu maqolaning maqsadi web scraping texnologiyalarining qanday ishlashi, ularning imkoniyatlari, cheklantirmalari va oqlanish uchun zarur shartlarini O'zbek tilida to'liq o'rganish. Maqolada web scraping-ning tarix, asosiy ta'rifi, texnik asoslari, turli vositalar, qo'llanilish sohalari va xavflari keltiriladi. Shuningdek, web scraping-ni to'g'ri va shaffof ravishda qo'llash bo'yicha tavsiyalar va eng yaxshi amaliyotlar ham ko'rib chiqiladi.

## **2. Web Scraping-ning Ta'rifi, Texnik Asoslari va Usullari**

**2.1. Web Scraping-ning Ta'rifi va Tarixiy Rivoji:** Web scraping veb-saytlardan avtomatik ravishda ma'lumotlarni ekstraktatsiya qiluvchi texnologiya. Bu jarayonda HTTP so'rovlar veb-serverga yuboriladi, HTML javoblar qabul qilinadi va keyin toza ma'lumotlar ekstraktatsiya qilinadi. Web scraping-ning ildizlari 1990-yillarning oxirida veb-saytlarning paydo bo'lishi bilan bog'liq. Dastlab web scraping faqat kitob o'qish indexlari uchun qo'llanilgan, ammo bugun bozor monitoring, tanlama tadqiqotlari, ma'lumot tahlili va boshqa ko'plab sohalarda keng qo'llanilmoqda.

**2.2. Web Scraping Turlarini va Usullari:** Web scraping-ni quyidagicha ajratib bo'ladi: (1) Statik web scraping - HTML saytlardan to'g'ridan-to'g'ri ma'lumotlarni ekstraktatsiya qilish; (2) Dinamik web scraping - JavaScript orqali yuklanadigan kontentni qabul qilish; (3) API-asosida yig'ish - to'g'ridan-to'g'ri REST yoki GraphQL API-lar orqali ma'lumotlarni olish. Shuningdek, web scraping usullarini HTML parsing, DOM manipulation, headless browsing va session management orqali amalga oshiriladi.

### **2.3. Web Scraping Asosiy Vositalar va Kutubxonalar**

<b>Dasturlash Tili</b>	<b>Kutubxona</b>	<b>Qo'llaniladigan Joylar</b>	<b>Xususiyatlari</b>	<b>Qiyinlik Darajasi</b>
<b>Python</b>	BeautifulSoup	HTML/XML parsing	Sodda va tez	Past
<b>Python</b>	Selenium	Dinamik kontentlar	JavaScript qo'llash	O'rta
<b>Python</b>	Scrapy	Katta loyihalar	Framework	Yuqori
<b>JavaScript</b>	Puppeteer	Headless Chrome	Screenshot va PDF	O'rta
<b>JavaScript</b>	Cheerio	Node.js parsing	jQuery-ga o'xshash	Past
<b>Java</b>	Jsoup	HTML parsing	Oxshash BeautifulSoup	O'rta

**Jadval 1.** Web scraping-ning asosiy vositalar va kutubxonalarini

## **3. Web Scraping-ning Qo'llanilish Sohalari va Faydalari**

**3.1. Bozor Monitoring va Raqobatchil Tahlili:** Web scraping raqobatchilarning narxlarini, mahsulot kataloglarini va marketing strategiyalarini kuzatib turish uchun qo'llaniladi . E-tijorat kompaniyalari bozor qiyofasini real vaqtda kuzatib, o'z narxlarini yanada samarali belgilab oladi . Bu esa konsumentlarga qancha yaxshi foyda beradi .

**3.2. Ma'lumot Tahlili va Tadqiqot:** Tibbiy, ijtimoiy va iqtisodiy tadqiqotlari uchun katta ma'lumot to'plamlari zarur . Web scraping sanoat raporti, ilmiy maqola abstraktlari, mamlakat statistikasi va ijtimoiy media ma'lumotlarini qidiruv qilish uchun qo'llaniladi . Bu ma'lumotlar keyin statistik analiz va mashina o'qitish modellari uchun qo'llaniladi .

**3.3. Mulk Qo'riqlab Turish va Ijtimoiy Monitoring:** Web scraping brands va brendlarning onlayn taqdimotini kuzatish uchun ishlatiladi . Qayta ishlangan asarlar, suvsatmalar va mualliflik huquqlari buzilishlari avtomatik ravishda aniqlash mumkin . Shuningdek, turli veb-tarmoqlarda firmalarning reputatsiyasi va izohlar kuzatib turiladi .

**4. Web Scraping-ning Xavflari, Etika va Huquq Masalalari**

**4.1. Huquq va Qonuni Masalalar:** Web scraping-ni qo'llashda veb-saytning shartlari va qonuni normalarni ta'minlash kerak . Kompyuter Fraudu va Ziyolatning Oldini Olish Akti (CFAA) kabi qonunlar, veb-saytdan ruxsatsiz ma'lumotlarni yig'ishni jinoyat deb belgilash mumkin . Alyoshnev Ommaviy Litsenziya va GDPR kabi ma'lumot xavfsizligi qonunlari ham web scraping-ni jiddiy cheklavchi qoidalarni kiritadi . Veb-sayt egasidan ruxsat olish yoki robots.txt fayl orqali ijozat olish web scraping-ni qonuni va etik qiladi .

**4.2. Texnik Xavflar va Ma'lumot Xavfsizligi:** Web scraping jarayonida IP adreslari bloklangan bo'lishi mumkin . Ayrim veb-saytlar bot-deteksiya tizimlari (reCAPTCHA, rate limiting) o'rnatgan . Xushomad (honeypot) safarbarliklar web scraper-larni qaytalash uchun tayyor bo'ladi . Shuningdek, API-lar orqali jo'natiladigan ma'lumotlar shifrlanmagan bo'lsa, man-in-the-middle hujumi muhim xavf bo'ladi .

**4.3. Web Scraping-ni To'g'ri Qo'llash Bo'yicha Tavsiyalar:** Web scraping-ni etik va qonuni qilish uchun quyidagi amaliyotlar talab qilinadi : (1) robots.txt faylini o'qish va hurmatlovchi; (2) veb-sayt egasidan ruxsat olish; (3) API-ni qo'llanish, agar mavjud bo'lsa; (4) veb-sayt serveriga ortiqcha yuk bermaydi (rate limiting); (5) foydalanuvchi-agent sarlavhasini to'g'ri o'rnatish; (6) shaxsiy ma'lumotlarni qo'ldan tashqari o'rnatmaydi; (7) ma'lumotlar xavfsizligini himoya qilish .

**5. Web Scraping-ni Mashin O'qitish va Kechinuvchi Analiz bilan Birlashtirish**

Web scraping-dan olingan ma'lumotlarni mashin o'qitish modellari uchun tayyorlash bugungi raqamli analizning asosiy masalasi . BeautifulSoup yordamida HTML-dan ekstraktatsiya qilingan ma'lumotlar pandas DataFrame-ga o'tkaziladi . Keyin ma'lumotlar tozalash (cleaning), normalizatsiya va feature engineering jarayonlardan o'tadi . Misol uchun, e-tijorat saytlaridan yig'ilgan mahsulot narxlari, sharh va reyting ma'lumotlari narx prognozlash modellari (regression) uchun qo'llanilishi mumkin . Shuningdek, ijtimoiy media ma'lumotlarini sentiment analizi va text mining orqali tahlil qilish uchun web scraping juda muhim . Veb-saytlardan yig'ilgan tarixiy ma'lumotlar time-series analizi uchun ham qo'llaniladi .

**Jadval 2. Web Scraping-ning Amaliy Qo'llanilish Misollari**

Soha	Qo'llanilish	Ma'lumotlarTuri	Foyda
E-tijorat	Narx monitoring	Mahsulot narxlari	Raqobat analizi
Ilm-fan	Tadqiqot ma'lumotlari	Maqolalar, abstraktlar	Meta-tahlil
Ijtimoiy media	Sentiment analizi	Post-lar, commentlar	Brand monitoring
Moliya	Saham qiymatlar	Real vaqt ma'lumotlar	Investitsiya qaror
Real estate	Mulk ma'lumotlari	Qiymat, lokatsiya	Bozor tahlili
Sog'liqni saqlash	Dorilar ma'lumoti	Narx, mavjudiyat	Tib xizmatlari

**Jadval 2.** Web scraping-ning turli sohalardagi amaliy qo'llanilishi

**6. Xulosa va Kelajkdagi Yo'nalishlar**

Veb-ma'lumotlarni avtomatik yig'ish (web scraping) bugungi raqamli transformatsiya davrida juda muhim texnologiya . Web scraping-ning to'g'ri qo'llanilishi ma'lumot tahlili, bozor monitoring, ilmiy tadqiqotlar va boshqa ko'plab sohalarda katta samaraga ega . Biroq, ushbu texnologiyaning xavflari va etika masalalari jiddiy ravishda ko'rib chiqilishi kerak . Robots.txt-ni hurmatlovchi, veb-sayt egasidan ruxsat olgan va rate limiting qo'llovchi web scraping to'g'ri amaliyot . Kelajakda yapay intellekt va mashin o'qitish texnologiyalari rivojlanishi bilan, web scraping-ning imkoniyatlari yanada ko'payishi kutilmoqda . Ammo, data privacy va cybersecurity muhimligi tobora oshib boriladi . Shuning uchun, web scraping-ni faqat etik va qonuni asoslar bilan qo'llash zarurdir .

### **Adabiyotlar**

- Qodirov, F. "OPTIMUM SOLUTIONS FOR THE DEVELOPMENT OF MEDICAL SERVICES IN PRIVATE CLINICS". Muhammad al-Xorazmiy nomidagi Toshkent Axborot Texnologiyalari Universiteti Qarshi filiali. (2022).
- Qodirov, F. "QR-KOD TEXNOLOGIYASI ASOSIDA ELEKTRON KUTUBXONA TIZIMINI DASTURIY VA APPARAT TAMINOTINI YARATISH". Muhammad al-Xorazmiy nomidagi Toshkent Axborot Texnologiyalari Universiteti Qarshi filiali. (2021).
- Qodirov, F. E., O. D. Doniyorov, and H. Shokirov Sh. "Basic Concepts Of Information Security In Information Systems. Wide Threats And Their Consequences." КОНЦЕПЦИИ УСТОЙЧИВОГО РАЗВИТИЯ НАУКИ В СОВРЕМЕННЫХ УСЛОВИЯХ (2021).
- Bozorova, Irina Jumanazarovna, and Dilduzaxon Mamasharipovna Karayeva. "Modern programming technologies and their role." интеллектуальный капитал ххi века. 2020.
- Kodirov, F. E., and J. E. Nematov. "BASIC TECHNOLOGY AND SERVICE MANAGEMENT MULTISERVICE NETWORKS." Инновации в технологиях и образовании: сб. ст. участников XII Между (2019).
- Sharipova S.B. "Veb-dasturlarining xavfsizligi va ma'lumot muhofazasi". O'zbekiston: Raqamli texnologiyalar, 2023.
- Qodirov F.E., Karayeva D.M. "Avtomatik ma'lumot yig'ishning texnik asoslari". Toshkent: Ilm va texnologiya, 2023.
- Nematov J.E. "Python va veb-parsing kutubxonalarini". O'zbekiston: Fan-ta'lim, 2022.
- Doniyorov O.D. "Ma'lumot xavfsizligi va raqamli etika". Toshkent: Ta'lim-dars, 2023.
- Ergashev R.T. "Web scraping-ni etik va qonuni qo'llash". O'zbekiston: Raqamli transformatsiya, 2023.