

WEB SCRAPING TEXNOLOGIYASI VA API DAN FOYDALANISH IMKONIYATLARI

Abduvaxobov Davlatjon Farxodjon o'g'li

FarDU Axborot tizimlari va texnologiyalari yo'nalishi 3-kurs talabasi

abduvaxobovdavlatjon@icloud.com

Sobirjonov Behzod Qahramonovich

FarDU Axborot texnologiyalari kafedrası Katta o'qituvchisi

bekzodbekqahromonovich@gmail.com

<https://doi.org/10.5281/zenodo.20354981>

Annotatsiya: Ushbu maqolada Web Scraping texnologiyasi va API (Application Programming Interface) dan foydalanish imkoniyatlari, ularning ishlash tamoyillari hamda zamonaviy axborot tizimlaridagi ahamiyati batafsil yoritilgan. Xususan, web scraping texnologiyasining internet sahifalaridan avtomatik ravishda ma'lumotlarni yig'ish, tahlil qilish va qayta ishlashdagi o'rni chuqur tahlil qilingan. Turli dasturlash vositalari va kutubxonalar yordamida ma'lumotlarni ajratib olish usullari, scraping jarayonining afzalliklari va amaliy qo'llanilish sohalari keng yoritilgan. Shuningdek, API texnologiyasining dasturlararo o'zaro aloqa o'rnatish, real vaqt rejimida ma'lumot almashish hamda tizimlarni integratsiyalashdagi imkoniyatlari ko'rib chiqilgan. REST API va boshqa zamonaviy API xizmatlarining ishlash prinsiplari, xavfsizlik masalalari hamda samaradorligi alohida tahlil etilgan. Bundan tashqari, web scraping va API'dan foydalanishda yuzaga keladigan muammolar, jumladan ma'lumotlar xavfsizligi, maxfiylik, saytlarning himoya mexanizmlari, server yuklamasi hamda huquqiy cheklovlar kabi masalalar ham muhokama qilingan. Tadqiqot natijalari web scraping va API texnologiyalari katta hajmdagi ma'lumotlarni tezkor yig'ish, tahlil qilish va samarali boshqarishda muhim vosita ekanligini ko'rsatadi. [1:25] [2:1]

Kalit so'zlar: Web Scraping, API, REST API, ma'lumotlarni yig'ish texnologiyasi, avtomatlashtirilgan tahlil, internet ma'lumotlari, JSON, HTML parsing, Python, ma'lumotlar integratsiyasi. [1:25] [2:1]

WEB SCRAPING TECHNOLOGY AND THE POSSIBILITIES OF USING APIs

Abduvaxobov Davlatjon Farxodjon o'g'li

3rd-year student, Information Systems and Technologies Department, FarDU

abduvaxobovdavlatjon@icloud.com

Sobirjonov Behzod Qahramonovich

Senior Lecturer, Department of Information Technologies, FarDU

bekzodbekqahromonovich@gmail.com

Abstract: This article provides a detailed overview of Web Scraping technology and the possibilities of using APIs (Application Programming Interfaces), including their operating principles and significance in modern information systems. In particular, the role of web scraping technology in automatically collecting, analyzing, and processing data from web pages is thoroughly examined. Methods of extracting data using various programming tools and libraries, as well as the advantages and practical applications of the scraping process, are widely discussed. Furthermore, the article explores the capabilities of API technology in establishing interaction between software systems, enabling real-time data exchange, and integrating different systems. The working principles, security issues, and efficiency of REST APIs and other modern API services are analyzed

separately. In addition, challenges encountered when using web scraping and APIs, including data security, privacy, website protection mechanisms, server load, and legal restrictions, are also discussed. The research findings demonstrate that web scraping and API technologies are important tools for rapidly collecting, analyzing, and efficiently managing large volumes of data. [1:25] [2:1]

Keywords: Web Scraping, API, REST API, data collection technology, automated analysis, internet data, JSON, HTML parsing, Python, data integration. [1:25] [2:1]

ТЕХНОЛОГИЯ WEB SCRAPING И ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ API

Абдувахобов Давлатжон Фарходжон ўғли

Студент 3 курса направления «Информационные системы и технологии»,
ФерГУ

abduvaxobovdavlatjon@icloud.com

Собиржонов Бехзод Кахрамонович

Старший преподаватель кафедры информационных технологий ФерГУ
bekzodbekqahromonovich@gmail.com

Аннотация: В данной статье подробно рассматриваются технологии Web Scraping и возможности использования API (Application Programming Interface), их принципы работы, а также значение в современных информационных системах. В частности, глубоко проанализирована роль технологии web scraping в автоматическом сборе, анализе и обработке данных с интернет-страниц. Широко освещены методы извлечения данных с использованием различных программных средств и библиотек, а также преимущества и практические области применения процесса scraping. Кроме того, рассмотрены возможности API-технологий в организации взаимодействия между программными системами, обеспечении обмена данными в режиме реального времени и интеграции различных систем. Отдельно проанализированы принципы работы REST API и других современных API-сервисов, вопросы безопасности и их эффективность. Также обсуждаются проблемы, возникающие при использовании web scraping и API, включая безопасность данных, конфиденциальность, механизмы защиты сайтов, нагрузку на сервер и правовые ограничения. Результаты исследования показывают, что технологии web scraping и API являются важными инструментами для быстрого сбора, анализа и эффективного управления большими объемами данных. [1:25.] [2:1.]

Ключевые слова: Web Scraping, API, REST API, технология сбора данных, автоматизированный анализ, интернет-данные, JSON, HTML parsing, Python, интеграция данных. [1:25.] [2:1.]

Kirish. Hozirgi raqamli davrda internet tarmog'ida mavjud ma'lumotlar hajmi keskin ortib bormoqda. Turli veb-saytlar, ijtimoiy tarmoqlar, onlayn xizmatlar va axborot platformalarida katta miqdordagi ma'lumotlar doimiy ravishda yangilanib boradi. Ushbu ma'lumotlardan samarali foydalanish, ularni yig'ish, qayta ishlash va tahlil qilish zamonaviy axborot texnologiyalarining muhim yo'nalishlaridan biriga aylangan. Shu sababli Web Scraping texnologiyasi hamda API xizmatlari katta ahamiyat kasb etmoqda. Web Scraping texnologiyasi internet sahifalaridan avtomatik ravishda kerakli ma'lumotlarni ajratib olish imkonini beruvchi zamonaviy usullardan biri hisoblanadi. Ushbu texnologiya yordamida foydalanuvchilar katta hajmdagi ma'lumotlarni

qisqa vaqt ichida yig'ish va ularni tahlil qilish imkoniyatiga ega bo'ladilar. Ayniqsa, marketing, moliya, ta'lim, sog'liqni saqlash va elektron tijorat sohalorida web scraping texnologiyasidan keng foydalanilmoqda. Mazkur texnologiya inson omilini kamaytirib, ma'lumotlarni yig'ish jarayonini avtomatlashtirish orqali samaradorlikni oshiradi. API (Application Programming Interface) esa turli dasturiy tizimlar o'rtasida o'zaro aloqa va ma'lumot almashinuvini ta'minlovchi muhim vosita hisoblanadi. API yordamida dasturlar real vaqt rejimida ma'lumot almashishi, tashqi xizmatlardan foydalanishi va tizimlarni integratsiyalashi mumkin. Hozirgi kunda REST API, SOAP API va GraphQL kabi texnologiyalar keng qo'llanilib, dasturiy ta'minot ishlab chiqish jarayonida muhim o'rin egallamoqda. [1:25.] [2:1.]

Web scraping va API texnologiyalari katta hajmdagi ma'lumotlar bilan ishlashda qulaylik yaratishiga qaramasdan, ularni qo'llash jarayonida ayrim muammolar ham yuzaga keladi. Jumladan, ma'lumotlar xavfsizligi, maxfiylik, server yuklamasi, veb-saytlarning himoya mexanizmlari va huquqiy cheklolvar muhim masalalardan biri hisoblanadi. Shu sababli ushbu texnologiyalardan samarali va qonuniy foydalanish zamonaviy axborot tizimlarida dolzarb vazifa bo'lib qolmoqda. [1:25.] [2:1.]

Mazkur maqolada Web Scraping texnologiyasi va API'dan foydalanish imkoniyatlari, ularning ishlash tamoyillari, afzalliklari hamda amaliy qo'llanilish sohalari tahlil qilinadi. Shuningdek, ushbu texnologiyalar bilan ishlash jarayonida uchraydigan muammolar va ularning yechimlari haqida ham fikr yuritiladi. [1:25.] [2:1.]

Natija va muhokama

Web scraping (veb-skreypping) va API (Application Programming Interface) ma'lumotlarni yig'ish va tahlil qilishning eng asosiy texnologiyalari hisoblanadi. Ular raqamli dunyoda turli xil ma'lumotlarni avtomatlashtirilgan tarzda olish imkonini beradi. Quyida ushbu texnologiyalarning imkoniyatlari, farqlari va qo'llanilishi bo'yicha batafsil ma'lumot keltirilgan. Web scraping — bu veb-sahifalardan avtomatlashtirilgan tarzda ma'lumotlarni olish texnikasi. Garchi ma'lumotlarni veb-sahifalardan foydalanuvchi tomonidan qo'lda olish mumkin bo'lgani bilan, odatda veb-skreyppingni nazarda tutganda, veb-resurslardan ma'lumotlarni avtomatlashtirilgan holatda GET so'rovlar orqali qabul qilishga aytiladi. Veb-sahifani veb-skreyppingni uni qabul qilishni va undan qirqib olishni o'z ichiga oladi. Qabul qilish - bu sahifani yuklash hisoblanadi. Yuklash va sahifalarni ko'rish ushbu texnologiyaning muhim tarkibiy qismlari bo'lib, ular ma'lumotlarni qidirishning ajralmas qismi hisoblanadi. Amaliy dasturlash interfeysi (inglizcha: Applications programming interface; qisqartmasi API) — ilova yoki operatsion tizim tomonidan tashqi dasturlar foydalanishi uchun taqdim etiladigan tayyor turkumlar, protseduralar, funksiyalar, tuzilmalar va konstantalar majmui. Foydalanuvchi interfeysidan farqli ravishda, ikki yoki undan ortiq kompyuter dasturlari o'zaro bog'lanishi va muloqot qila olishi uchun ishlatiladi. Bu turdagi aloqa yoki interfeysni ishlab chiqish, yoxud foydalanishni tavsiflovchi hujjat API spesifikasiyasi (inglizcha: API specification) deb nomlanadi.

Kompyuter va insonni bog'lovchi foydalanuvchi interfeysidan farqli ravishda, API kompyuterlar yoki dasturiy ta'minot qismlarini bir-biri bilan bog'lashda ishlatiladi. API oddiy foydalanuvchilar tomonidan emas, balki, dasturchilar tomonidan foydalanish uchun mo'ljallangan. [2:1.]

APIning asosiy maqsadlaridan biri tizimning ichki detallarini yashirib, dasturchi uchun kerak bo'ladigan qismlarini ochiq qilish va ichki detallar o'zgargan taqdirda ham ularni izchil saqlashdan iboratdir. API ma'lum bir tizim juftligi uchun moslashtirilgan, yoki, ko'plab tizimlar o'rtasida o'zaro ishlash imkonini beruvchi umumiy standart tarzida ishlab chiqilgan bo'lishi

mumkin.APIlarning bir qancha, jumladan, dasturlash tillari, dasturiy ta’minot kutubxonalari, operatsion tizimlar va kompyuter hardveyri uchun ishlatilnadigan turlari mavjud. API atamasi 1960-1970-yillarda paydo bo’lgan bo’lsa ham, ushbu texnologiyaning tarixi 1940-yillarga borib taqaladi. API atamasining hozirgi kundagi qo’llanishi odatda internetga ulangan kompyuterlar o’rtasida aloqa o’rnatish imkonini beruvchi veb-APIlarga ishora qiladi.APIlarning rivojlanishi mikroservislarining mashhurlikka erishishiga ham sababchi bo’lgan.Ma’lumotlarni yig’ish va tahlil qilish zamonaviy axborot texnologiyalarining asosiy yo’nalishlaridan biri bo’lib, internetdagi katta hajmdagi ma’lumotlarni avtomatlashtirilgan tarzda yig’ish, qayta ishlash va integratsiya qilishni o’z ichiga oladi. Bu jarayonda Python dasturlash tili eng ommabop vosita hisoblanadi.Internetdan ma’lumotlarni avtomatlashtirilgan tarzda yig’ish web scraping yoki web crawling deb ataladi. Ular veb-sahifalardagi ma’lumotlarni (matn, narxlar, rasmlar) strukturaviy formatga o’tkazadi. Agar ma’lumot rasmiy API orqali taqdim etilsa, undan foydalanish tavsiya etiladi. Agar ma’lumot ochiq bo’lsa-yu, lekin API bo’lmasa, unda web scraping texnologiyalari eng yaxshi tanlovdir. [2:1.]

Xulosa: Web Scraping va API texnologiyalari zamonaviy axborot tizimlarida ma’lumotlarni yig’ish, qayta ishlash va tahlil qilishning muhim vositalaridan biri hisoblanadi. Ushbu texnologiyalar internetdagi katta hajmdagi ma’lumotlarni avtomatlashtirilgan tarzda olish imkonini berib, turli sohalarda samaradorlikni oshirishga xizmat qiladi. Web scraping texnologiyasi veb-sahifalardan kerakli ma’lumotlarni avtomatik ravishda ajratib olish imkonini bersa, API esa dasturlar va tizimlar o’rtasida xavfsiz hamda qulay ma’lumot almashuvini ta’minlaydi.Tadqiqot davomida web scraping va API texnologiyalarining ishlash prinsiplari, afzalliklari hamda qo’llanilish sohalari tahlil qilindi. Web scraping texnologiyasi ochiq veb-resurslardan ma’lumot yig’ishda samarali vosita ekanligi, API esa rasmiy va ishonchli ma’lumot almashish usuli sifatida keng qo’llanilishi aniqlandi. Ayniqsa, Python dasturlash tili ushbu texnologiyalar bilan ishlashda qulay va samarali vosita sifatida muhim o’rin egallaydi. Shuningdek, ushbu texnologiyalarni qo’llash jarayonida ma’lumotlar xavfsizligi, maxfiylik, server yuklamasi va huquqiy cheklovlar kabi muammolar mavjudligi ham ko’rib chiqildi. Shu sababli web scraping va API dan foydalanishda qonuniy me’yorlarga rioya qilish, saytlarning foydalanish siyosatini hisobga olish hamda ma’lumotlarni himoyalash muhim ahamiyatga ega.Umuman olganda, Web Scraping va API texnologiyalari raqamli iqtisodiyot va zamonaviy dasturiy tizimlarning ajralmas qismiga aylanib bormoqda. Ular yordamida katta hajmdagi ma’lumotlarni tezkor yig’ish, tahlil qilish va integratsiya qilish imkoniyati yaratiladi hamda bu esa turli tashkilotlar va foydalanuvchilar uchun samarali qarorlar qabul qilishga xizmat qiladi. [1:25.] [2:1.]

Adabiyotlar, References, Литературы:

- 1.Axborot texnologiyalari atamalarining izohli lug’ati 2023, s. 25.
2. Reddy, Martin. API Design for C++. Elsevier Science, 2011 — 1-bet. ISBN 9780123850041. [2:1.]
3. Lane. Intro to APIs: History of APIs” (en-US). Postman (2019-yil 10-oktyabr). — When you hear the acronym “API” or its expanded version “Application Programming Interface,” it is almost always in reference to our modern approach, in that we use HTTP to provide access to machine readable data in a JSON or XML format, often simply referred to as “web APIs.” APIs have been around almost as long as computing, but modern web APIs began taking shape in the early 2000s.”. Qaraldi: 2020-yil 18-sentyabr.[2:1.]
4. Wood. „Global Cloud Microservices Market (2021 to 2026)” (en-US). businesswire.com (2021-yil 25-avgust). Qaraldi: 2022-yil 29-mart.

5. Python for Data Analysis / McKinney W. *Python for Data Analysis*. — USA: O’Reilly Media, 2022.
6. Web Scraping with Python / Mitchell R. *Web Scraping with Python: Collecting More Data from the Modern Web*. — USA: O’Reilly Media, 2024. [1:25.]
7. RESTful Web APIs / Richardson L. Amundsen M. *RESTful Web APIs*. — USA: O Reilly Media, 2013. [2:1.]
8. Mozilla Foundation. API va HTTP texnologiyalari bo’yicha ma’lumotlar. MDN Web Docs [2:1]
9. Fielding R. *Architectural Styles and the Design of Network-based Software Architectures*. — Doctoral dissertation, University of California, 2000.