

“TELEGRAMM” IJTIMOY TARMOG‘I RESURSLARIDAN MATNLI MA’LUMOTLARNI YIG‘ISH

Jaxongir Dadajon-o‘g‘li RISQALIYEV

Katta o‘qituvchi

O‘zbekiston Respublikasi Ichki ishlar vazirligi

Malaka oshirish instituti

Toshkent, O‘zbekiston

jaxongir.tmi93@gmail.com

ORCID: 0009-0009-1202-5911

Annotatsiya

Ushbu maqolada “Telegramm” ijtimoiy tarmog‘idagi kanallar va guruhlardan ma’lumotlarni yig‘ish, ular tarkibidan matnli ma’lumotlarni ajratib olish usullari tahlil qilingan.

Tayanch so‘zlar: “Telegramm”, matnli ma’lumotlarni yig‘ish, tahlil qilish.

СБОР ТЕКСТОВЫХ ДАННЫХ ИЗ РЕСУРСОВ СОЦИАЛЬНОЙ СЕТИ “TELEGRAMM”

Жахангир Дадажан-угли РИСКАЛИЕВ

Старший преподаватель

Институт повышения квалификации

Министерство внутренних дел Республики Узбекистан

Ташкент, Узбекистан

jaxongir.tmi93@gmail.com

ORCID: 0009-0009-1202-5911

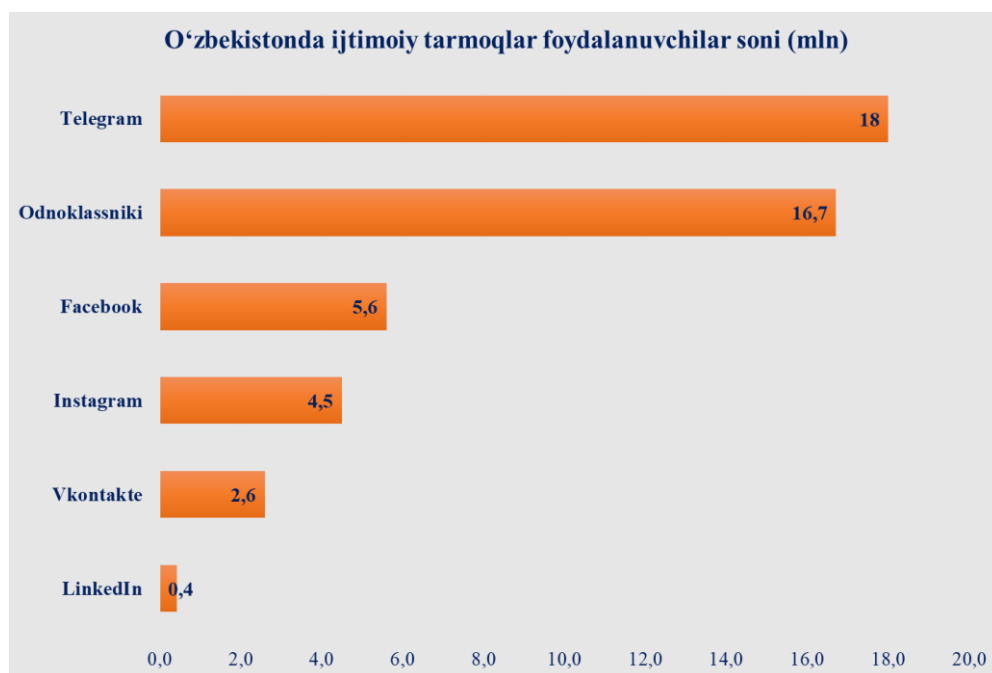
Аннотация

В данной статье анализируются методы сбора данных из каналов и групп в социальной сети “Telegramm”, а также извлечения текстовых данных из содержания собранных данных.

Ключевые слова: “Telegramm”, сбор текстовых данных, анализ.

Hozirgi kunda odamlar do‘stlari, oila a‘zolari va turli guruhlar bilan muloqot qilishlari uchun ijtimoiy tarmoqlardan foydalanib kelishmoqdalar. Tashkilot va korxonlar o‘z mahsulotlarini reklama qilish, sotish, mijozlarning talab-takliflarini o‘rganib borishni ijtimoiy tarmoqlar orqali amalga oshiradilar. O‘zbekiston aholisining ham turli xil ijtimoiy tarmoqlardan foydalanishi rivojlanib bormoqda.

Tahliliy ma'lumotlarga ko'ra, "Telegramm" ijtimoiy tarmog'i O'zbekistonda ommabop ijtimoiy tarmoq bo'lib turibdi. O'zbekiston "Telegramm" kanallari soni va ularning auditoriyasi bo'yicha dunyoda ikkinchi (Rossiyadan keyin) o'rinda turadi. "Telegramm"da 18 million o'zbekistonlik o'z akkauntlarini yaratgan. Keyingi o'rinlarda "Odnoklassniki"da 16,7 million, "Facebook" (Meta)da 5,6 million, "Instagram"da 4,5 million, "VKontakte"da 2,6 million, "LinkedIn"da – 432 ming va "Twitter"da 51,6 ming o'zbekistonlik foydalanuvchi o'z akkauntiga ega [150].



1-rasm. O'zbekistonda ijtimoiy tarmoqlar foydalanuvchilar soni.

"Telegramm" – bu foydalanuvchilarga real vaqt rejimida muloqot qilish imkonini beruvchi va tobora ommalashib borayotgan xabar almashish ilovasi. Butun dunyo bo'ylab "Telegramm" foydalanuvchilar soni 2023-yil yakunlariga ko'ra 800 milliondan oshdi [14]. "Telegramm"da foydalanuvchilar matnli xabar, rasm, audio, video, stiker va hajmi 2 GB gacha bo'lgan fayllarni yuborishi mumkin.

"Telegramm" ijtimoiy tarmog'i ommalashib borayotganing sababi foydalanuvchi maxfiyligiga e'tibor berganligidadir. "Telegramm"ning maxfiylikka qaratilgan yondashuvi tufayli undan noqonuniy faoliyatlarda foydalanishiga ham olib keldi. Indoneziyadagi terrorchilar radikal mafkuralarni tarqatish va zo'raonlik

hujumlarini amalga oshirish uchun ko'rsatmalarni berib borishda "Telegramm" ijtimoiy tarmog'idan foydalanganlar [13]. "Telegramm" orqali ekstremistlar o'z mafkuralarini tarqatishda kanal va guruhlardan foydalangan [9]. Bundan tashqari, pornografik kontentni tarqatish uchun ham ishlatilmoqda [10]. Bu "Telegramm"ni ko'p noma'lum yashirin tomonlarini tadqiq qilishga arziydi. M.Nasari va H.Zamani beshta axborot agentligining rasmiy "Telegramm" kanallarida joylashtirilgan yangiliklarni to'plash orqali "Telegramm" platformasida yangiliklarning tarqalishini o'rgandi [4;2]. N.Prucha, A.S.Yayla va A.Speckhard, A.Shehabat, T.Mitew va Y.Alzoubi tadqiqot ishlarida IShID (Iroq va Shom Islom davlati) kabi terroristik tashkilotlar "Telegramm"dan kontent va mafkurani tarqatish, shuningdek, jangchilar va terrorchilarni yollash kabi turli maqsadlarda qanday foydalanishini o'rganishgan [5, 6, 8]. M.Taghipour, F.Aboutorabi, V.Zarrabi va H.Asghari "Telegramm" guruh va kanallaridan fors tilidagi ma'lumotlarni yig'ish va tahlil qilish imkoniyatiga ega dasturiy vositani ishlab chiqishgan [7;2]. Dunyo bo'ylab ijtimoiy tarmoqlarni tadqiq qilish rivojlanib bormoqda, ammo ijtimoiy tarmoqlardan o'zbek tilidagi matnli ma'lumotni yig'ish va tahlil qilish bo'yicha juda kam ish qilingan.

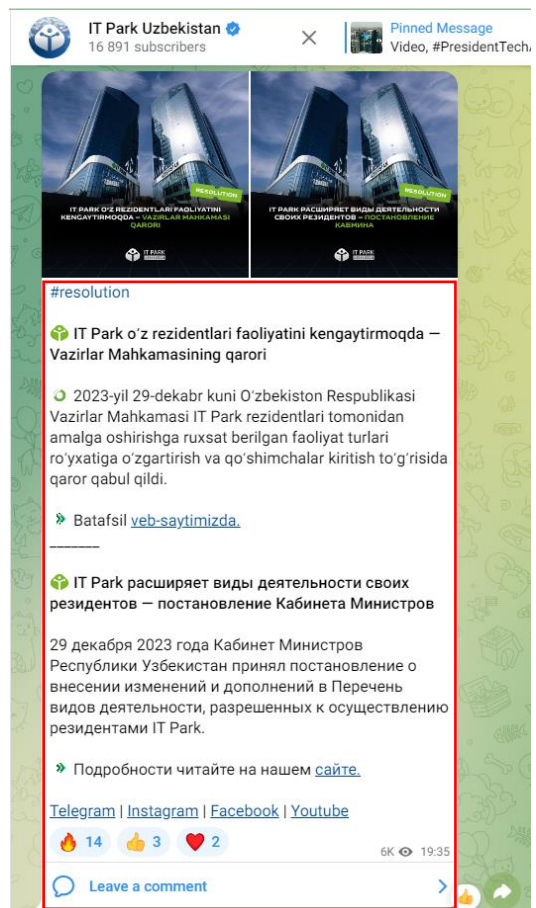
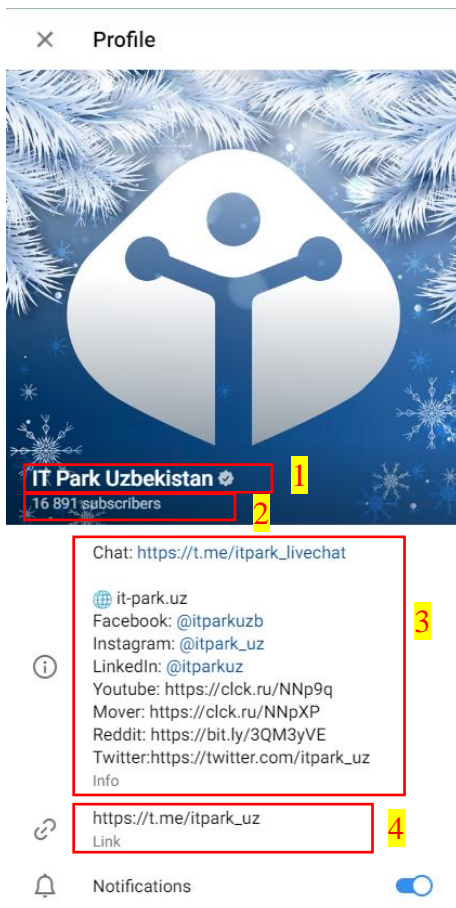
"Telegramm" foydalanuvchilari kontentni boshqa foydalanuvchilar bilan bo'lish uchun kanallar, umumiy yoki shaxsiy guruhlar yaratishi mumkin.

Kanallar: ommaviy xabarlarini katta auditoriyaga yetkazish uchun vositadir. "Telegramm" kanallari cheksiz miqdordagi obunachilarga ega bo'lishi mumkin va kanalga kontent faqat kanal administratori tomonidan joylashtiriladi. Kanaldagi xabarlarini "Telegramm" ijtimoiy tarmog'idagi har kim o'qishi mumkin yoki kanal admini tomonidan cheklov o'rnatilgan bo'lsa, faqat kanalga a'zo bo'lganlarga ko'ra oladi. Kanal admini ruxsat bergan bo'lsa, foydalanuvchilar joylashtirilgan kontentga izoh bildirishi mumkin. Kanal o'zining nomi, obunachilar soni, ma'lumot va noyob kanal havolasiga ega bo'lishi mumkin. 2-rasmda "Telegramm" kanalga misol keltirilgan. "Telegramm" kanallari "Telegramm" guruhlaridan farqli ravishda, kanallar xabarlar yonida – ularni joylashtirgan shaxs o'rniga kanal nomi va fotosuratini ko'rsatadi. Kanal administratori kanalga joylangan postlarga izohlar

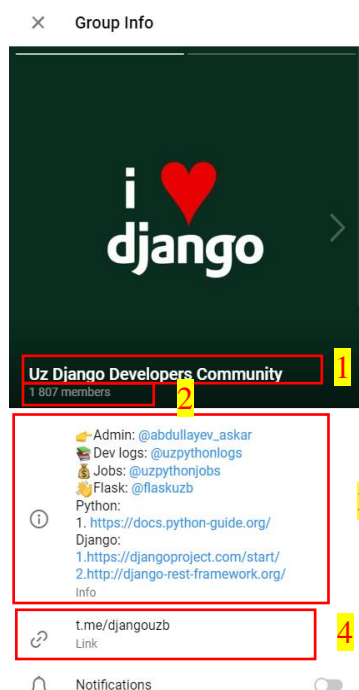
qoldirish uchun ochiq qoldirish ham mumkin. Kanaldagi har bir post ko‘rish sonini aniqlash uchun hisoblagichiga egadir. Kanaldagi obunachilar ro‘yxati faqat kanal egasi va administratorlariga ko‘rinadi. Obunachilar kanaldan kim ro‘yxatdan o‘tgan va qaysi akkaunt (administrator tomonidan tayinlangan qo‘shimcha administrator)lar kanalni boshqarayotganini ko‘ra olmaydi. Jamoat arboblari va tashkilotlari o‘zlarining ommaviy kanallarini tasdiqlashlari mumkin, shunda foydalanuvchilar uning rasmiy ekanligini bilishadi. Tasdiqlanganlik belgisi ko‘k nishon orqali ko‘rinadi, hamda kanal nomini yonida aks etadi. “Telegramm”da kanal, guruh yoki botni tasdiqlash uchun @VerifyBot havolasiga xabar yuboriladi.

Guruhlar: 200 mingtagacha a‘zolar bir-birlari bilan xabar almashish mumkin bo‘lgan kuchli vositadir. “Telegramm”da guruhlar shaxsiy yoki ommaviy bo‘lishi mumkin. Shaxsiy guruhlar hamma uchun ochiq emas va ularni ijtimoiy tarmoq resurslari ichidan qidirish orqali topib bo‘lmaydi. Shaxsiy guruhlariga qo‘shilishning yagona yo‘li bu guruh adminidan taklif havolasini olishdir. Bunday taklif havolalari *t.me* bilan boshlanadi. Agar guruh ochiq bo‘lsa, bunday umumiy guruhlar har kimga ko‘rinadi. Guruh qoidalarini buzgan foydalanuvchini guruh admini chiqarib tashlashi mumkin. Guruh o‘zining nomi, obunachilar soni, guuh haqida ma’lumot va noyob kanal havolasiga ega bo‘lishi mumkin. 3-rasmda “Telegramm” umumiy guruhiga misol keltirilgan.

Kanallar va guruhlardagi baham ko‘rilgan matnli ma’lumotlarni yig‘ish va to‘plangan ma’lumotlarni tahlil qilish qimmatli axborotlarni taqdim etishi mumkin.

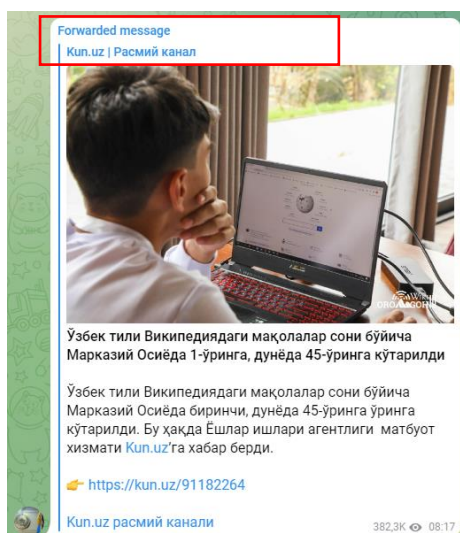


2-rasm. Telegram kanaliga misol: a) kanal haqidagi asosiy ma'lumotlar: nomi (1), obunachilar soni (2), ma'lumot (3) va kanal havolasi (4); b) kanal tomonidan chop etilgan xabar.



3-rasm. "Telegramm" guruhiga misol: a) guruh ma'lumotlari: nomi (1), obunachilar soni (2), ma'lumot (3) va guruh havolasi (4); b) guruhda foydalanuvchilar tomonidan qoldirilgan postlar.

“Telegramm”da guruhlar va kanallar o‘rtasida to‘g‘ridan-to‘g‘ri bog‘liqlik bo‘lmasada, yo‘naltirish havolasi orqali bog‘lanishi mumkin (4-rasm). “Telegramm” kanallari va guruhlaridan matnli ma’lumotlarni yig‘ish va tahlil qilishga qiziqish ortib bormoqda. Biroq, “Telegramm” ijtimoiy tarmog‘idan foydalanuvchilarning shaxsiy chatlaridagi ma’lumotlarni o‘qish, yig‘ish va ulardan foydalanish mumkin emas. “Telegramm” ijtimoiy tarmog‘idagi katta ma’lumotlar oqimi tufayli ma’lumotlarni yig‘ish va tahlil qilish qiyin vazifadir. Bunga O‘zbekistonda 150 mingdan ortiq “Telegramm” kanallari va 26 mingga yaqin guruhlarining mavjudligi, “Telegramm” kanallarining umumiy auditoriyasi 956 740 000 kishini tashkil etayotgani va foydalanuvchilar tomonidan har kuni joylashtiriladigan juda ko‘p ma’lumotlar sabab bo‘ladi [12].



4-rasm. “Telegramm” guruhida yo‘naltirilgan xabar.

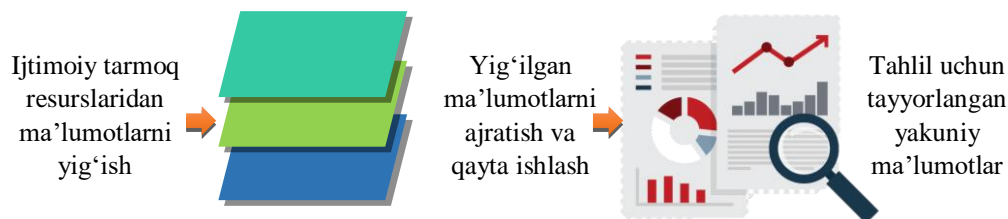
Ma’lumotlarni kanallar va guruhlardan qo‘lda yig‘ish usuli ham mavjud. Biroq, bu ko‘p vaqt talab qiladi va katta hajmdagi ma’lumotlar uchun samarali usul hisoblanmaydi. N.Manaco va D.Arnaudo tadqiqotchilar uchun ijtimoiy tarmoq resurslaridan ma’lumotlarni yig‘ishni 3 ta usulini ko‘rib, solishtirib chiqishgan [3;9]. Bular:

- tayyor uchinchi tomon dasturiy vositalaridan foydalanish;
- to‘g‘ridan-to‘g‘ri API (Application Programming Interface)dan foydalanish;
- veb qirqish (web scraping).

API yordamida ma’lumotlarni olish ko‘p hollarda qonuniydir, chunki API ma’lumotlari ijtimoiy tarmoqlar tomonidan tartibga solinadi va foydalanuvchilar

huquqlarini buzmaydigan qilib tuziladi. Veb qirqish esa ko‘p hollarda xizmat ko‘rsatish shartlarini buzish hisoblanib, noqonuniy deb baholanadi.

Ijtimoiy tarmoq resurslaridagi ma’lumotlar odatda strukturalanmagan bo‘ladi. Bu ma’lumotlarni asosan matn, rasm, audio yoki video bo‘lishi mumkin. Shuning uchun faqat kerakli bo‘lgan matnli ma’lumotlarni ajratib olish lozim. Har xil turdagi ma’lumotlarni tahlil qilish uchun turli xil usullar mavjuddir. Ijtimoiy



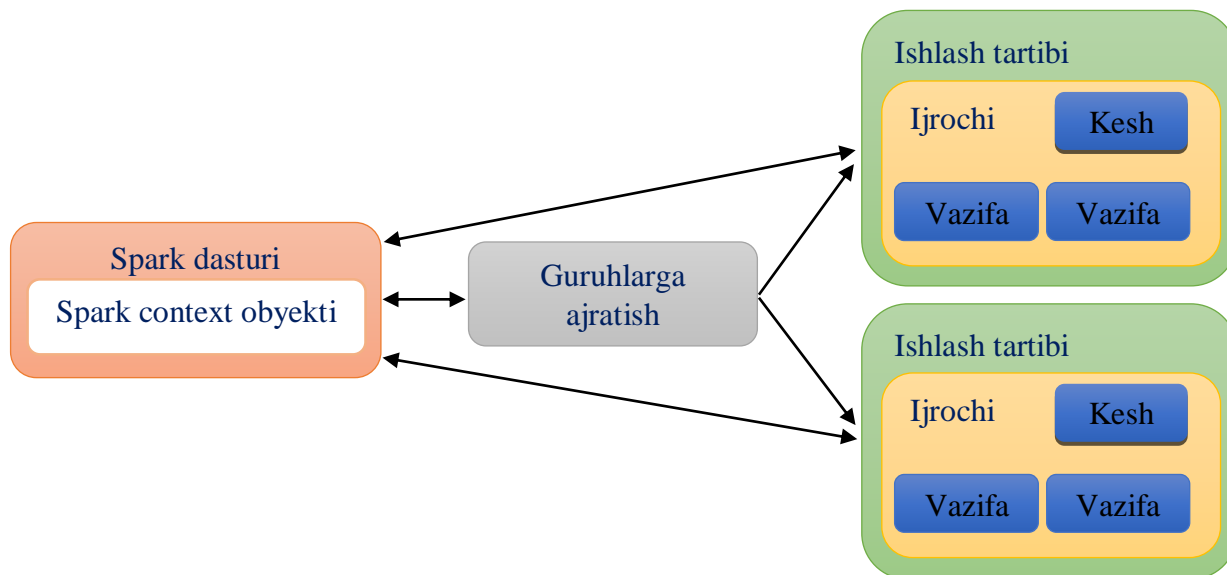
1-sxema. Strukturalanmagan ma’lumotlarni bilan ishlash jarayoni.

tarmoq resurslaridagi ma’lumotlar bilan ishlash jarayoni 1-sxemada keltirilgan.

Matnli ma’lumotlarning ham hajmi katta bo‘lishi mumkin. Shuning uchun yig‘ish bosqichini o‘zida, qiziqish uyg‘otadigan ma’lumotlarni ma’lum bir kalit so‘zlar yordamida filtrlash lozimdir. Buning uchun hozirgi kunda keng qo‘llanilayotgan, real vaqtda katta ma’lumotlarni qayta ishlash xususiyatiga ega bo‘lgan ochiq manbali “Apache Spark” dasturiy vositasi qo‘llaniladi. “Apache Spark” dasturidan foydalanilganda, xotiradan ma’lumotlarni o‘qish, amallarni bajarish va natijalarni saqlash uchun birgina qadam kerak bo‘ladi, bu esa ishni bajarilishini ancha tezlashtiradi. 2-sxemada “Apache Spark” tizimi ishlash jarayonining arxitekturasini berilgan.

“Telegramm” ijtimoiy tarmog‘i resurslaridan to‘g‘ridan-to‘g‘ri matnli ma’lumotlarni yig‘ish imkoniyati mavjud emas. Biroq, mutaxassislar “Telegramm” kanallari va guruhlaridan matnli ma’lumotlarni ma’lum vaqt oralig‘ida ajratib olish imkoniyatini ega TG-API (“Telegramm” API) va Python dasturida yordamida qilingan dasturiy vositani ishlab chiqishdi [11]. J.Baumgartner, S.Zannetou, M.Squire va J.Blackburnlar esa 27,8 mingdan ortiq kanal va 2,2 million foydalanuvchilarning 317 million xabarlaridan iborat Pushshift “Telegramm” ma’lumotlar to‘plamini taqdim qilishgan [1;845]. M.L.Morgia, A.MeI va

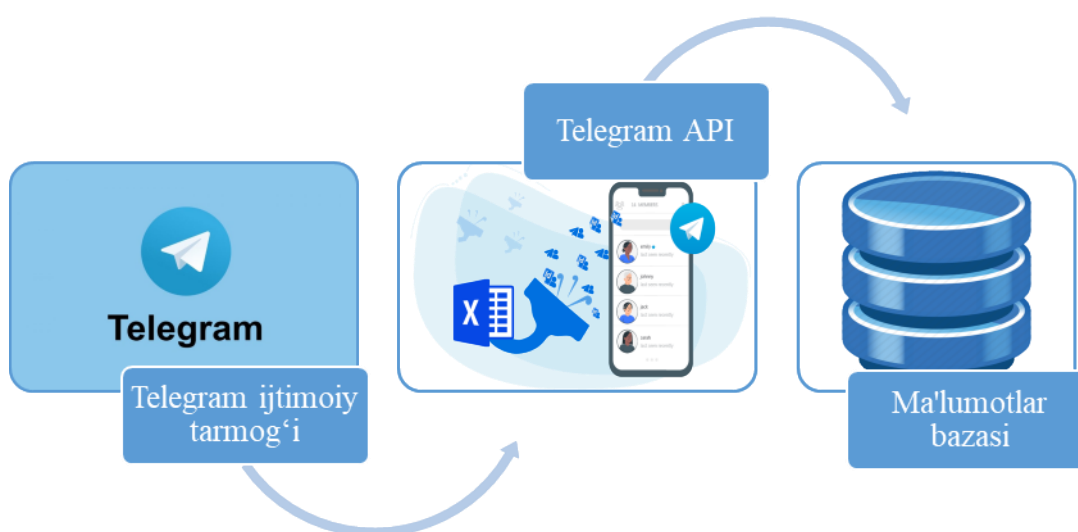
A.M.Mongardinilar esa 120 979 ta “Telegramm” kanallari va 400 milliondan ortiq xabarlarni o‘z ichiga olgan TGDataset ma’lumotlar to‘plamini taklif qilishgan [2;1]. Kanallar va guruhlardagi ma’lumotlarni ma’lum vaqt bo‘yicha ajratib olish esa olingan ma’lumotlarni tahlil qilishning samarali usuli hisoblanadi.



2-sxema. Apache Spark tizimi ishlash jarayonining arxitekturasi.

O‘zbekistonga oid “Telegramm” guruh va kanallar ro‘yxatini statistik ma’lumotlarni taqdim etuvchi mashhur “Tgstat” xizmatidan olish mumkin [12]. “Tgstat” eng mashhur kanallar, kanal uchun obunachi soni, mavzu toifasi, o‘shish sur‘ati yoki faol guruhlar, mashhur postlar kabi ma’lumotlarni taqdim etadi. “Telegramm” uchun API kalitlar <https://core.telegram.org/api> saytidan olinadi. “Telegramm” API yordamida Python dasturining “telethon” va “pandas” kutubxonalari yordamida “Telegramm” kanallari va guruhlaridan yakka tartibda yoki ommaviy ravishda matnli ma’lumotlar ajratib olinadi, so‘ngra olingan ma’lumotlarni o‘z ichiga olgan JSON fayllari yaratiladi. Rasm, audio, video va matn fayllar fayl kengaytmasi pdf, docx, xlsx bo‘lgan fayllarga saqlanadi. Ajratib olingan foydalanuvchi postlariga qo‘shimcha ravishda kanal/guruh haqidagi ma’lumotlar ham saqlanadi. Bundan tashqari, JSON faylini CSV fayliga o‘tkazish va CSV fayli yordamida natijalarni tahlil qilish qulayroqdir. “Telegramm” ijtimoiy tarmog‘i guruh va kanallaridan matnli ma’lumotlarni yig‘ish jarayoni 3-sxemada berilgan.

“Telegramm” messenjeri oxirgi yillarda sezilarli mashhurlikka erishdi. Shuning uchun ijtimoiy tarmoqda bo‘layotgan faoliyatni o‘rganish va tahlil qilish juda muhimdir. Bu tahlilchilar yoki reklama beruvchilar uchun foydali bo‘ladi. Kelajakda o‘zbek tilidagi matnli ma’lumotlarni yig‘ish, yig‘ilgan matnli ma’lumotlarni NLP (Natural language processing) va mashinaviy o‘qitish (machine learning) texnologiyalarini qo‘llagan holda intellektual tahlillash uchun dasturiy vositalarni ishlab chiqish rejalashtirilgan.



3-sxema. Matnli ma’lumotlarni yig‘ish jarayoni.

FOYDALANILGAN ADABIYOTLAR:

1. Baumgartner, J., Zannetou, S., Squire, M. and Blackburn, J. 2020. The Pushshift Telegram Dataset. Proceedings of the Fourteenth International AAAI Conference on Web and Social Media. <https://ojs.aaai.org/index.php/ICWSM/article/view/7348/7202>
2. Massimo La Morgia, Alessandro Mei, and Alberto Maria Mongardini. 2023. TGDataset: a Collection of Over One Hundred Thousand Telegram Channels. https://www.researchgate.net/publication/369116144_TGDataset_a_Collection_of_Over_One_Hundred_Thousand_Telegram_Channels

3. Monaco, N., Arnaudo, D. 2020. Data analytics for social media monitoring. <https://www.ndi.org/publications/data-analytics-social-media-monitoring>
4. Naseri, M., Zamani, H. 2019. Analyzing and Predicting News Popularity in an Instant Messaging Service. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 1053–1056.
5. Prucha, N. 2016. Is and the jihadist information highway projecting influence and religious identity via telegram. Perspectives on Terrorism 10(6).
6. Shehabat, A., Mitew, T. and Alzoubi, Y. 2017. Encrypted jihad: Investigating the role of telegram app in lone wolf attacks in the west. Journal of Strategic Security 10(3):27–53.
7. Taghipour, M., Aboutorabi, F., Zarrabi, V. and Asghari, H. 2018. An Integrated Text Mining Platform for Monitoring of Persian News Agencies. https://scholar.google.com/citations?view_op=view_citation&hl=zh-CN&user=QYuFV28AAAAJ&citation_for_view=QYuFV28AAAAJ:2osOgNQ5qMEC
8. Yayla, A.S., and Speckhard, A. 2017. Telegram: The mighty application that isis loves. International Center for the Study of Violent Extremism. https://www.researchgate.net/publication/316789344_Telegram_the_Mighty_Application_that_ISIS_Loves
9. <https://english.alarabiya.net/News/saudi-arabia/2023/01/04/Saudi-Arabia-works-with-Telegram-to-remove-15-021-951-extremist-content-in-2022>
10. <https://ipeeworld.com/fix-this-channel-cannot-be-displayed-issue-in-telegram/>
11. <https://os2int.com/toolbox/batch-scraping-from-telegram-channels-and-groups-using-tg-api/>
12. <https://uz.tgstat.com/>
13. <https://www.bbc.com/news/business-40627739>
14. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

15. <https://yuz.uz/ru/news/internet-i-sotsialne-seti---neotemlemaya-chast-nashey-jizni?view=tsifrovoy-obraz-jizni>