# BEAUTIFULSOUP ПАКЕТІНІҢ КӨМЕГІМЕН ДЕРЕКТЕРДІ ТАЛДАУҒА АРНАЛҒАН АҚПАРАТТЫ ЖИНАУ

**Турсынгалиева Г.Н., Жанат А.**
*Карагандинский университет имени академика Е.А. Букетова, Караганда, Казахстан*
E-mail: gulim_tursyngali@mail.ru, janatavdulazim@mail.ru

Кез келген үлкен деректерді талдаудың мақсаты – зерттелетін жағдайды толық түсіну, тенденцияларды, оның ішінде жоспардан теріс ауытқуларды анықтау, болжау және ұсыныстаралу. Осы мақсатқа жету үшін деректерді талдаудың келесі міндеттері қойылады:

- ақпарат жинау,
- ақпаратты құрылымдау,
- заңдылықтарды анықтау, талдау,
- болжау және ұсыныстар алу.

Талдау жасап, әдемі графиктерді құрастырмас бұрын, ақпаратты жинау алғашқы сатылардың бірі екендігі анық.

Жұмыстың мақсаты – COVID-19 короновирусының таралуы жайлы болжамдарды талдау үшін қажетті деректерді интернет көзінен BeautifulSoup пакеті арқылы алу.

Кәзіргі таңда веб-сайттардан ақпаратты жинаудың түрлі тәсілдері өтек көп. Солардың бірі – сайтты парсингтеу. Бұл веб-сайттардан ақпаратты алу әдісі және бірінші кезекте құрылымдалмаған деректерді - HTML пішіміндегі - вебтегі құрылымдық деректерге: дерекқорларға немесе электрондық кестелерге түрлендіруге бағытталған. Веб-сайтты парсингтеу HTTP арқылы немесе веб-шолғыш арқылы Интернетке тікелей кіруді қамтиды. Веб-деректерді шығара алатын әртүрлі тілдерде көптеген кітапханалар мен фреймворктар бар болса да, Python көптеген веб-скрапинг мүмкіндіктеріне байланысты көп қолданылады.

Қажетті ақпаратты жинау үшін Python-ды пайдаланып, келесі әрекеттерді орындалады:

- Деректерді шығарғымыз келетін беттің URL мекенжайын алу;
- Беттің HTML мазмұнын көшіру немесе жүктеу;
- HTML мазмұнын талдау және қажетті деректерді алу.

Бұл реттілік қалаған беттің URL-мекенжайына өтуге, HTML мазмұнын алуға және қажетті деректерді талдауға көмектеседі. Бірақ кейде деректерді алу үшін алдымен сайтқа кіріп, содан кейін белгілі бір мекенжайға өту керек. Бұл жағдайда сайтқа кіру үшін тағы бір қадам қосылады.

HTML мазмұнын талдау және қажетті деректерді алу үшін BeautifulSoup кітапханасы пайдаланылды. Бұл - HTML және XML құжаттарын парсингтеуге арналған Python пакеті. Ал, дұрыс парсингтеу үшін сайттардың құрылымын түсіну керек. Олардың барлығы дерлік HTML тілі арқылы жасалған.

Қазіргі таңда COVID-19 короновирусының таралуы жайлы болжамдар өте көп жасалуда. Бұл болжамдарды жасамас бұрын, қажетті ақпаратты интернет көзінен жинау қажет. Алдымен, болжам мен талдау үшін инфекциялар, өлім және сауығулар туралы тарихи деректер керек. COVID-19 вирусы туралы деректер Worldometer ақпараттық сайтында (https://www.worldometers.info/coronavirus/countries-where-coronavirus-has-spread/ ) еркін қол жетімді.

Ақпаратты сайттан алмас бұрын Python программалау ортасына «requests», «bs4» и «texttable кітапханалары орнатылды.

Қажетті кітапханаларды орнатқан соң, келесі қадам - BeautifulSoup пакеті арқылы қажетті кодты жазып, әртүрлі елдердегі жаңа коронавирустың (COVID-19) расталған, өлім-жітім, сауығып кеткен және белсенді жағдайлары туралы соңғы деректерді але үшін, ең алдымен, файлдың жоғарғы жағында requests мен BeautifulSoup кітапханалары импортталады. Бұдан кейін url айнымалысы ақпарат келетін беттің мекенжайын сақтайды. Бұл айнымалы мән requests.get() функциясына жіберіледі де, нәтиже жауап айнымалысына

тағайындалады. Әрі қарай, жауап мәтінін soup айнымалысына орналастыру үшін BeautifulSoup() конструкторын қолданылды. Пішім ретінде lxml таңдалынып, айнымалы шығарылды.

Деректерді оқи алатын форматта көрсету үшін texttable кітапханасын қолданылды.

Әр мемлекеттің короновирус жайлы ақпараты жүктелді. Осы ақпаратты пайдаланып әрі қарай талдаулар жасап, оны визуалдауға болады.

Нәтижесінде BeautifulSoup requests қосымшасының көмегімен жұмыс істелініп, қажетті сайттарға html сұраулары ұйымдастырылды. BeautifulSoup арқылы алынған ақпарат өңделініп, Қазақстандағы короновирус бойынша статистика, болжамдар жасалынды. Парсерлер интернеттен беттерді жүктеп алып, оларды құрамдас бөліктерге талдап берді: тақырып, сурет, мәтін... Оның көмегімен сайттан гигабайттық пайдалы ақпарат жүктелініп алынды.

BeautifulSoup - Python-дағы бірнеше скрапингке арналған кітапханалардың бірі. Онымен жұмысты істеу өте жеңіл болғандықтан, сценарийлерді интернеттен деректерді жинау және құрастыру үшін пайдалануға болады, ал нәтиже деректерді талдау және басқа сценарийлер үшін пайдаланылуы мүмкін.

**Қолданылған әдебиеттер тізімі**


1. Программирование на Python : научное издание / М. МакГрат. - М. : ЭКСМО, 2020. - 192 с.
2. Аубакиров, Г. Д. Языки программирования Python: учеб. пособие для организаций техн. и проф. образования / Г. Д. Аубакиров, А. Г. Хмыров. - 2-е изд. - Астана : Фолиант, 2011. - 203 с.
3. Лутц,М. Программирование на Python. Т.2/М.Лутц. –М.:Символ,2016.- 992 с
4. https://python-scripts.com/beautifulsoup-html-parsing
5. https://dvmn.org/encyclopedia/modules/bs4-tutorial/


## ИСПОЛЬЗОВАНИЕ OLAP-ТЕХНОЛОГИЙ В БИЗНЕСЕ


**Шаяхметова Б.К.[1], Омарова Ш.Е.[2], ДроздВ.Г.[2], УлаковН.С.[2], Есмагамбетов Т.У.[2]**

*1 Карагандинский университет им.Е. А.Букетова, Караганда, Казахстан*

*2 Карагандинский университет Казпотребсоюза, Караганда, Казахстан*

E-mail: sheo_1953@mail.ru,vgdrozd@mail.ru, nazar-123@mail.ru, timur198300@mail.ru


Современные социально-экономические условия трансформации общества определили стремительный переход к новой информационной ступени развития, обусловленной проникновением информационных технологий во все сферы человеческой деятельности.

Формирование и развитие информационного общества, его гуманизация нашли свое отражение в различных сферах и прежде всего, в экономических информационных системах.

Развитие промышленных предприятий, связанных с созданием и реализацией продукции и услуг, невозможно без использования информационных технологий. В настоящее время очень много информациипубликовано о OLAP технологиях. Информационныесистемы масштаба предприятия содержат приложения предназначенные для комплексного многомерного анализа данных, их динамики, тенденций и т.д.Такой анализ в конечном итоге призван содействоватьпринятию решений. Часто такие эти системы называются системами поддержки принятия решений.

Системы поддержки принятия решений обычнообладают средствами предоставления пользователюагрегатных данных для различных выборок из исходного набора в удобном виде для восприятия и анализа. Как правило, такие агрегатные функции составляют многомерный (не реляционный) набор данных,оси которого содержат параметры, а ячейки – зависящие от них агрегатные данные – причем хранитьсятакие данные могут и в реляционных таблицах[1].

Реализацией этих положений является многомерное представление информации в специальных базах данных OLAP и доступ к таким базам обеспечивается через клиентское приложение. Основным объектом OLAP-технологий и баз OLAP является куб - информация, сохраненная в специальном формате многомерных данных.