**Практична робота № 1.**

**(2 години)**

**Тема:** ***Імпорт (парсинг) даних із сайтів засобами електронних таблиць.***

**Мета:** Сформувати практичні навички формування інформаційних баз використовуючи технологію парсингу та аналізу даних засобами елетронних таблиць.

**Зміст роботи.**

1. Використовуючи хмарний сервіс Google Spreadsheets створити таблицю з іменем в якому відобразити тему та прізвище автора.
2. На першому листі таблиці проілюструвати роботу функцій: IMPORTRANGE(), ImportHTML(), ImportXML(), IMAGE(), GOOGLETRANSLATE(). До кожного результату додати коментар з поясненнями.
3. На листі 2 створеної таблиці відобразити результат парсингу сайту Житомирської політехніки отримавши інформацію про розклад вашої групи поточного тижня.
4. Ознайомитись з процесом парсингу даних із сайтів до таблиць Excel за допомогою додатку ***Power Query*** (Презентація лекції)***.***
5. На Листі 1 електронної таблиці отримати дані з будь якого сайту, що розміщені в таблиці. На першому рядку відобразити назву та посилання на сайт джерело.
6. На Листі 2 електронної таблиці отримати інформацію з нетабличних даних про товар та його ціну. Сайт та товар вибрати на свій розсуд.
7. Отримані таблиці відформатувати засобами MS Excel для подальшого аналізу. В заголовку таблиць відобразити назву джерела та посилання на нього.

**Вимоги до результату роботи.**

На перевірку надсилається посилання на таблицю Google Spreadsheets, файл MS Excel в якій відображено результати парсингу та звіт створений в текстовому редакторі в якому засобами скролінгу відобразити основні етапи парсингу даних а саме:

* + - отримання даних в *Power Query* ;
    - накладені фільтри;
    - заміна значень (один приклад);
    - розділення стовпців (один приклад)
    - результат *Power Query;*
    - результат *в* MS Excel до форматування.

**!!!! Важливо.**

***Power Query працює починаючи з 16 версії офісу.***

Тип даних необхідно задавати в *Power Query.*

В *Power Query* також можна використовувати функції. Наприклад, якщо дані в результаті будуть відображені в 1 стовпчик, можна скористатися функцією List.Split() для формування списків по елементам та функцією Table.FromRows() для перетворення рядків в колонки.

**Контрольні запитання.**

1. Яка мета імпорту та парсингу даних сайтів.
2. Як створюються файли в Google сервісах?
3. В чому полягає призначення функцій IMPORTRANGE(), ImportHTML(), ImportXML(), IMAGE(), GOOGLETRANSLATE().
4. Що є параметрами функцій IMPORTRANGE(), ImportHTML(), ImportXML(), IMAGE(), GOOGLETRANSLATE().
5. Як отримати програмний код елементу сайту?
6. Як отримати адресу XPath?
7. Як отримати порядковий номер таблиці для відображення парсингованих даних?
8. Як можна використати інформацію рядка: //\*[@id="idx-content"]/div[3]/table?
9. Як вставити зображення в клітину Google Spreadsheets?
10. Як отримати доступ до Power Query в MS Excel?
11. Назвіть основні функції Power Query.
12. Як отримати програмний код сайту в вікні Power Query?
13. Як скористатися фільтром і яка в ньому потреба?
14. Який функціонал Power Query ви використали в процесі роботи?
15. Які засоби форматування існують в Power Query і які ви використовували?
16. Як перемістити дані з Power Query в електронну таблицю?