МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ "КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ"

Факультет прикладної математики Кафедра програмного забезпечення комп'ютерних систем

Лабораторна робота № 1

з дисципліни "Бази даних 2. БД на основі XML" тема "Вивчення базових операцій обробки XML-документів"

Виконав	Зарахована
студент III курсу	""20p.
групи КП-83	викладачем
Коваль Андрій Олександрович (прізвище, ім'я, по батькові)	Петрашенко Андрієм Васильовичем (прізвище, ім'я, по батькові)

варіант № 8

Мета роботи:

Здобуття практичних навичок створення програм, орієнтованих на обробку XML-документів.

Варіант: 8

Базова сторінка (завдання 1)	Зміст завдання 2	Адреса інтернет-магазину (завдання 3)
www.golos.ua	Середня кількість	www.petmarket.ua
	текстових фрагментів	

Хід роботи:

Завдання 1

Основна обробка html відбувається у декількох функціях за допомогою xpath.

Далі, проходячись по усім стоірінкам з даних посилань, отримуємо зображення та текстові елементи. Нажаль, не вдалося отримати дані елементи без прямого втручання через TypeScript. Можливо, проблема у використаній бібілотеці 'xpath' для JS.

```
Отримання зображень та тексту з документу
export const getImagesAndTexts = (doc: Document, url: string): PageData => {
 // @ts-ignore
 // xpath doesn't work, cannot exclude script and style =( ?????
 // */text()[not(ancestor::script)]
 // */text()[not(parent::script or style)]
 // */text()[not(parent::script | style)]
 // *[not(self::script | style)]/text()
 const nodes = selectHTML("//x:img/@src | //x:*/text()", doc)
      .filter((node: Node) =>
      // @ts-ignore
      !IGNORED TAGS.includes((node.parentNode as Node)?.tagName)
 );
 const fragments: Fragment[] = nodes.map((node: Node & WithNodeValue) => ({
      type: node.nodeName === 'src' ? 'image' : 'text' as 'image' | 'text',
      data: node.nodeValue,
 })).filter(fragment => !!fragment.data.trim());
 return {
      url,
      fragments,
 };
```

Результуючий файл task1.xml вийшов досить великий за розміром (2492 рядки)

```
    task1.xml 

    ★
                                                                          ზე •o → →
out > 6 task1.xml >  data >  page
       You, 2 days ago | 1 author (You)
       <?xml version="1.0"?>
      <data>
   3 > <page url="https://www.golos.ua/">--
         </page>
  410 > <page url="https://www.golos.ua/news">--
         </page>
 416 > <page url="https://www.golos.ua/publikatsii">--
         </page>
 422 > <page url="https://www.golos.ua/intervyu"> --
 428 > <page url="https://www.golos.ua/anonsy">--
         </page>
 434 > <page url="https://www.golos.ua/blogs">--
         </page>
 440 > <page url="https://www.golos.ua/dose">--
 445
         </page>
 446 > <page url="https://www.golos.ua/archive">--
         </page>
  778 > <page url="https://www.golos.ua/gallery">--
 912 > <page url="https://www.golos.ua/video">--
         </page>
 1140 > <page url="https://www.facebook.com/golosmedia">--
         </page>
 1328 > <page url="https://t.me/golosua">--
 1347
         </page>
 1348
        <page url="https://twitter.com/golosua">
          <fragment type="image">https://abs.twimg.com/errors/logo46x38.png</fragment>
1350
         <fragment type="text">This browser is no longer supported.//fragment>
           <fragment type="text">...
          </fragment>
           <fragment type="text">Help Center</fragment>
           <fragment type="text">Terms of Service</fragment>
           <fragment type="text">Privacy Policy</fragment>
           <fragment type="text">Cookie Policy</fragment>
          <fragment type="text">Imprint</fragment>
1358
          <fragment type="text">Ads info</fragment>
          <fragment type="text">
1360
           © 2021 Twitter, Inc.
          </fragment>
1362
         </page>
         <page url="https://www.youtube.com/user/tvgolosua/feed">...
 1379
         </page>
 1380 >
        <page url="https://golos.ua/feed/">...
 1446
         </page>
1447 > <page url="https://www.golos.ua/category/politika">--
1714
         </page>
1715 > <page url="https://www.golos.ua/category/ekonomika">---
         </page>
1972 > <page url="https://www.golos.ua/category/obshhestvo">--
2236
2237 > <page url="https://www.golos.ua/category/duhovnost">--
2491 /page>
 2492 </data>
```

Рис 1. task1.xml, вигляд у редакторі

Основний метод для завдання 1 export const run = async (url: string): Promise<string> => { const pagesData = await getPagesDataRecursive(url); const xml = pagesDataToXML(pagesData); await fsPromise.writeFile(config.TASK1_FILE_NAME, xml); return xml; Використані типи даних export interface Fragment { type: 'text' | 'image', data: string; export interface PageData { url: string; fragments: Fragment[], export interface WithNodeValue { nodeValue: string; value: string;

Завлання 2

Усі обчислення відбувалися за допомогою xpath

```
pages: pagesCount,
    textFragments: textFragmentsCount,
    avgTextFragments: avgTextFragmentsCount,
};
};
```

Вивід програми

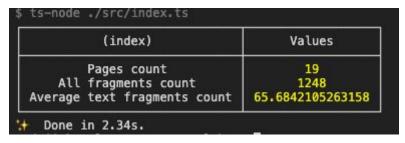


Рис 2. Вивід завдання 2

Завдання 3

Основним завдання ϵ діставання продуктів із зазначеного у завдання магазину за допомогою web scraping. За допомогою храth та мінімального втручання з боку мови програмування було стврорено відповідну функцію

```
get-products.ts
export const getProducts = (doc: Document, maxCount?: number): Product[] => {
 const images: Node[] = selectHTML("//x:img[contains(@class,
'catalogCard-img')]/@src", doc) as unknown as Node[];
 // //x:*[@class='catalogCard-title']/text() this didn't work(
 const titles: Node[] = selectHTML("//x:*[@class='catalogCard-title']/*[1]",
doc) as unknown as Node[];
 const prices: Node[] =
selectHTML("//x:div[@class='catalogCard-price']/text()", doc) as unknown as
 const count = Math.min(images.length, titles.length, prices.length, maxCount);
 const products: Product[] = new Array(count).fill(0).map((_, idx) => ({
      img: images[idx].nodeValue,
      name: titles[idx].firstChild.nodeValue.trim(),
      price: prices[idx].nodeValue.trim(),
 }));
 return products;
```

Основна логіка цього завдання є досить простою.

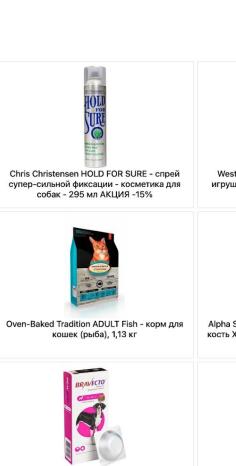
Функція run для виконання 3 завдання export const run = async (url: string): Promise<string> => { const html = await getHtml(url); const doc = HTMLToXMLDocument(html); const products = getProducts(doc, PRODUCTS_TO_SCRAP_COUNT); const xml = productsToXML(products); await fsPromise.writeFile(config.TASK3_FILE_NAME, xml); return xml; }

Завдання 4

На жаль не було знайдено робочої бібліотеки для трансформації xml в xhtml через xslt для JavaScript. Для виконання завдання було вирішено скористатися сервісом https://cedricvb.be/xml-xslt-to-xhtml-transform/. Завдяки ньому було виконано даже завдання.

```
XSLT для трансформації
```

```
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
      <link rel="stylesheet"</pre>
href="https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0/css/bootstrap.min.css"
integrity="sha384-Gn5384xqQ1aoWXA+058RXPxPg6fy4IWvTNh0E263XmFcJlSAwiGgFAW/dAiS6J
Xm" crossorigin="anonymous" />
      </head>
      <body class="container row text-center mx-auto">
      <xsl:for-each select="root/product">
         <div class="card m-1 p-2 col-5">
             <div class="card-image">
             <img>
             <xsl:attribute name="src">
                   https://petmarket.ua<xsl:value-of select="img" />
             </xsl:attribute>
             </img>
             </div>
             <h5 class="card-title">
                <xsl:value-of select="name" />
             </h5>
             <xsl:if test="@price">
                <xsl:value-of select="price" />
                </xsl:if>
         </div>
      </xsl:for-each>
   </body>
      </html>
  </xsl:template>
</xsl:stylesheet>
```



БРАВЕКТО XL - таблетка от блох и клещей для собак 40-56 кг %

Oven-Baked Tradition SALMON Grain Free влажный беззерновой корм для кошек (лосось) - 156 г АКЦИЯ!





Рис 3. Результуючий UI трансфорованого XHTML

Усі додаткові програмні модулі та реалізація можуть бути знайденими у GitHub peпозиторії https://github.com/ZioVio/DB-course/tree/master/term2/lab1 До опису винесено лише найважливіші програмні файли та функції, які становлять основну частину завдання роботи.

Висновки

Завдяки даній роботі я дізнався про XSLT, XPath, XHTML та здобув практичні навички створення програм, орієнтованих на обробку XML-документів.