

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ВАСИЛЯ СТУСА
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ І ПРИКЛАДНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ СИСТЕМ УПРАВЛІННЯ

ВАКУЛЕНКО ЮЛІЯ ОЛЕГІВНА

Допускається до захисту:
в.о. завідувача кафедри
інформаційних систем управління,
д-р екон. наук, професор
_____ Ольга АНІСІМОВА
« _____ » _____ 2021 р.

**ВПРОВАДЖЕННЯ ТА ПЕРСПЕКТИВИ ВИКОРИСТАННЯ
ПРОГРАМ-ПАРСЕРІВ В ЗАКЛАДАХ ВИЩОЇ ОСВІТИ**

Спеціальність 029 «Інформаційна, бібліотечна та архівна справа»

Кваліфікаційна (магістерська) робота

Науковий керівник:
Щербіна О.С., доцент кафедри
інформаційних систем управління,
канд. екон. наук, доцент

Оцінка: _____ / _____ / _____
(бали / за шкалою ЕКТС / за національною шкалою)

Голова ЕК: _____
(підпис)

Вінниця – 2021

Вакуленко Ю. О. Впровадження та перспективи використання програм-парсерів у закладах вищої освіти. Спеціальність 029 Інформаційна, бібліотечна та архівна справа, освітня програма «Документаційна та інформаційна підтримка управлінської діяльності». Донецький національний університет імені Василя Стуса, 2019. – 85 с.

У магістерській роботі розкрито сутність парсингу як автоматизованого методу пошуку інформації. Представлено та досліджено конкретний приклад пошукової машини Google Spreadsheet, запропоновано та визначено її перспективи впровадження у відділ кадрів Донецького національного університету імені Василя Стуса.

Ключові слова: програми-парсери, заклад вищої освіти, відділ кадрів, парсинг, інформація, комунікація, комунікаційний процес.

Табл. 1. Рис. 16. Бібліограф.: 84. найм.

Vakulenko Yuliia. Implementation and prospects of using parser programs in higher education institutions. Specialty 029 Information, Library and Archival Studies, Educational Program “Documentation and information support for management activities”. Vasyl’ Stus Donetsk National University, Vinnytsia, 2021.

In the master's thesis reveals the essence of parsing as an automated method of information retrieval. A specific example of the Google Spreadsheet search engine is presented and researched, its prospects for implementation in the personnel department of Vasyl Stus Donetsk National University are proposed and determined.

Keywords: parser programs, higher education institution, human resources department, parsing, information, communication, communication process.

Tab. 1. Fig. 16. Bibliograph.: 84 items.

ЗМІСТ

ВСТУП	4
РОЗДІЛ 1 ТЕОРЕТИЧНІ ЗАСАДИ ВПРОВАДЖЕННЯ ТА ВИКОРИСТАННЯ ПРОГРАМ-ПАРСЕРІВ	7
1.1 Сутність поняття «парсинг»	7
1.2 Етапи розвитку програм-парсерів в Україні	19
1.3 Мета, завдання та передумови впровадження програм-парсерів в закладах вищої освіти	30
Висновки до розділу 1	33
РОЗДІЛ 2 АНАЛІЗ ВИКОРИСТАННЯ ПРОГРАМ-ПАРСЕРІВ В ЗАКЛАДАХ ВИЩОЇ ОСВІТИ (НА ПРИКЛАДІ ДОНЕЦЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ ІМЕНІ ВАСИЛЯ СТУСА)	35
2.1 Аналіз комунікаційного процесу закладу вищої освіти	35
2.2 Система комунікацій у внутрішньому середовищі закладу вищої освіти	47
2.3 Механізми функціонування програм-парсерів у відділі кадрів Донецького національного університету імені Василя Стуса	53
Висновки до розділу 2	58
РОЗДІЛ 3 КОНЦЕПЦІЯ ФОРМУВАННЯ КОМУНІКАЦІЇ ВИКОРИСТАННЯМ ПРОГРАМ-ПАРСЕРІВ У ВІДДІЛІ КАДРІВ	59
3.1 Проблеми впровадження методу парсингу в закладах вищої освіти в напрямі забезпечення інформацією відділу кадрів	59
3.2 Перспективи використання програм-парсерів у відділі кадрів закладу вищої освіти	63
Висновки до розділу 3	74
ВИСНОВКИ	76
СПИСОК ВИКОРИСТАНИХ ПОСИЛАНЬ	78

ВСТУП

Актуальність теми дослідження. Нині основною та головною потребою людського життя є інформація та комунікація. Проте, особливо користувачем мережі інтернет, витрачається багато часу на пошук саме потрібної інформації, адже зазвичай для цього використовується великий масив даних. Виникає потреба в пришвидшенні пошуку інформації та отриманні результату, якого потребує користувач. Сьогодні програми-парсери проводять синтаксичний, лексичний, автоматизований аналіз, обробляють велику кількість інформаційних запитів, розбирають документи, вилучаючи звідти необхідні та корисні користувачу дані. Тому актуальність впровадження та подальші перспективи використання програм-парсерів не викликає сумнівів.

Мета дослідження. Метою магістерської роботи є розробити концепцію формування комунікації з використанням програм-парсерів у відділі кадрів закладу вищої освіти для удосконалення інформаційної діяльності.

Реалізація поставленої мети вимагає вирішення наступних **завдань**:

- розкрити сутність поняття «парсинг»;
- висвітлити етапи розвитку програм-парсерів в Україні;
- визначити мету, завдання та передумови впровадження програм-парсерів в закладах вищої освіти;
- провести аналіз комунікаційний процес закладу вищої освіти;
- проаналізувати систему комунікацій у внутрішньому середовищі закладу вищої освіти;
- проаналізувати механізми функціонування програм-парсерів у відділі кадрів Донецького національного університету імені Василя Стуса;
- виокремити проблеми впровадження методу парсингу в закладах вищої освіти в напрямі забезпечення інформацією відділу кадрів;
- означити перспективи використання програм-парсерів у відділі кадрів закладу вищої освіти.

Об'єктом дослідження є теоретичне й методичне забезпечення комунікації та прийняття ефективних управлінських рішень в діяльності закладів вищої освіти.

Предмет дослідження – процес впровадження та перспективи використання програм-парсерів в закладах вищої освіти.

Методи дослідження. Дане дослідження базується на структурно-функціональному підході щодо розкриття сутності поняття «парсинг», його мети, завдання, що пропонується як впровадження у відділ кадрів Донецького національного університету імені Василя Стуса. Характер роботи обумовив необхідність логіко-діалектичного методу пізнання, що дозволяє досліджувати проблему впровадження повністю. Метод системного аналізу відобразився у етапах функціонування програм-парсерів, а метод абстракції яскраво був показаний при наведенні прикладів наведеного дослідження.

Наукова новизна дослідження. Обґрунтовано необхідність впровадження та використання програм-парсерів у закладах вищої освіти. Проаналізовано комунікаційний процес та систему комунікацій у внутрішньому середовищі закладу вищої освіти. Виділено методи оцінювання ефективності роботи програм-парсерів у відділі кадрів Донецького національного університету імені Василя Стуса. Виокремлено проблеми впровадження методу парсингу в закладах вищої освіти в напрямі забезпечення інформацією відділу кадрів.

Практичне значення отриманих результатів. Рекомендації магістерської роботи можуть бути використані закладами вищої освіти в процесі впровадження та використання програм-парсерів.

Результати та окремі пропозиції проведеного наукового дослідження доведені до практичного рівня застосування (*довідка про використання результатів та окремих пропозицій дослідження відділом кадрів Донецького національного університету імені Василя Стуса*). Обґрунтовано необхідність використання парсингу в роботі відділу кадрів Донецького національного університету імені Василя Стуса. Використання програм-парсерів є

багатоплановим в кадровій діяльності та має перспективи стати головним процесом здійснення комунікації в закладі вищої освіти. Програми-парсери спрямовані на використання новітніх інформаційно-комунікаційних технологій, що допомагають вдосконалити управління, покращують взаємодію працівників та сприяють соціальним інноваціям.

Апробація результатів дослідження. Окремі аспекти магістерської роботи були представлені на VI Всеукраїнській науковій студентській конференції «Інформаційні технології і системи в документознавчій сфері» (м. Вінниця, 2021). Результати дослідження опубліковано у науковому журналі «Бібліотекознавство. Документознавство. Інформологія» (м. Київ, № 2, 2021).

Структура дослідження. Магістерська робота складається зі вступу, трьох розділів, що в свою чергу поділяються на підрозділи, висновків, списку використаних посилань. Загальний обсяг роботи становить 86 сторінок, з них 9 сторінок – список використаних посилань, що містить 84 позиції. Робота включає 1 таблицю та 16 рисунків.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ЗАСАДИ ВПРОВАДЖЕННЯ ТА ВИКОРИСТАННЯ ПРОГРАМ-ПАРСЕРІВ

1.1 Сутність поняття «парсинг»

В сучасному світі інформації ми часто стикаємося з проблемою, коли необхідно швидко вивантажити список товарів з інтернет-магазину або зібрати інформацію з сайту, при цьому зберігши певні дані, які нам необхідні. Наприклад, такі як назва, ціна, картинка, посилання на товар і ін. Можна уявити, скільки часу йде, на те, щоб зробити все це вручну![1].

Обробити вручну значні обсяги даних в мережі інтернет неможливо. В такому випадку задіюють спеціально розроблене програмне забезпечення – парсер. Ці програми створюються для різних завдань, перш за все пов'язаних з маркетингом і просуванням сайтів. Наприклад, компанія AVADA MEDIA пропонує як готові рішення в даній області, так і парсери на замовлення.

Парсер – програма, що забезпечує парсинг, тобто синтаксичний аналіз контенту в мережі до певної математичної моделі, створеної на одній з мов програмування, зазвичай на Python, PHP або Java.

Алгоритм роботи програми-парсера в кожному випадку приблизно однаковий:

- програма отримує доступ до ресурсу в мережі інтернет;
- завантаження коду сторінки, парсинг якою необхідно провести;
- читання і обробка інформації;
- надання результату в одному із зручних форматів – .html, .xml, .sql та інших. [1].

Програма порівнює заданий набір слів або значень зі знайденим в мережі – з урахуванням заданих же обмежень. Алгоритм подальших дій зі знайденим

контентом також прописаний заздалегідь – це може бути надання інформації або певні дії, наприклад, вчинення ставок при парсингу спортивних подій.

Парсинг - це свого роду імпортер даних. Скопіювати інформацію можна не тільки в інтернет-магазині, але і на сайтах спільних покупок, в групі в соціальних мереж, також якщо це стосується освітнього процесу. Головне, підібрати правильний інструмент і алгоритм.

За допомогою парсера вийде вивантажити найрізноманітнішу інформацію. Це може бути не тільки назва, а й опис, відео контент, зображення, основні характеристики і не тільки.

Парсер робить все самостійно. Ця обставина імпонує багатьом.

Парсинг - це трудомістка і монотонна робота. Набагато простіше довірити таку автоматизації. Це оптимальний вибір, коли стоїть завдання зробити ресурс успішним у довгостроковій перспективі. [2].

Парсинг пропонує інструментарій, який допомагає витягати потрібні значення з будь-яких форматів даних. Витягнуті дані зберігаються в окремому файлі - або на локальному комп'ютері, або в хмарі або на хостингу, або безпосередньо в базі даних. Це процес, який запускається автоматично.

Проаналізувати зібрану інформацію допомагає програмне забезпечення. Що значить парсити в програмуванні? Працюючий парсер надсилає запит типу GET на сайт-«донор», який повинен натомість віддати дані. В результаті цього запиту створюється HTML-документ, який буде проаналізовано програмою. Потім парсер здійснює в ньому пошук необхідних даних і виконує перетворення в потрібний формат. [2].

Існує 2 різних методи виконання веб-парсинга:

- доступ до сайту www через протоколи HTTP, HTTPS або веб-браузер;
- доступ з використанням бота.

Парсинг іноді плутають з краулінгом. Це відбувається тому, що дані процеси схожі. Обидва мають ідентичні варіанти використання. Різниця - в цілях.

Обидві операції «заточені» на обробку даних з сайтів. Процеси автоматизовані, оскільки використовують роботів-парсерів. Але це всього лише боти для обробки інформації або контенту. [3].

І парсинг, і краулінг проводять аналіз веб-сторінок і пошук даних за допомогою програмних засобів. Вони ніяк не змінюють інформацію, представлену там, і працюють з нею безпосередньо.

Парсинг збирає дані і сортує їх для видачі за заданими критеріями. І це необов'язково відбувається в Інтернеті, де роблять парсинг веб-сторінок. Йдеться про дані, а не про те, де вони зберігаються.

Наприклад, ви хочете попрацювати над ціновою аналітикою. Для цього ви запускаєте створені парсери товарів і цін на них, щоб зібрати інформацію з Avito або з будь-якого інтернет-магазину. Таким же чином можна аналізувати дані фондового ринку, оголошення з нерухомості і так далі.

Краулінг або веб-сканування - прерогатива пошукових роботів або павуків. Краулінг включає в себе перегляд сторінки цілком в пошуках інформації і її індексацію, включаючи останню букву і точку. Але ніякі дані при цьому не беруться. Інтернет-бот, він же - парсер пошукової системи - теж систематично переглядає всесвітню павутину для того, щоб знайти сайти і описати їх вміст. Найважливіша відмінність від краулера - він збирає дані і систематизує їх.

Те, що роблять Google, Яндекс або Yahoo - приклад веб-сканування. Це теж свого роду парсинг. Простими словами, це коли пошукові машини сканують сайти і використовують отриману інформацію для індексації.

Деякі не розуміють різниці: парсер або парсинг? Парсер - програмне рішення, а парсинг - процес.

Тобто програма для парсинга - це парсер. Вона призначена для автоматичної обробки та вилучення даних.

Для аналізу заданого тексту таке ПО зазвичай використовує окремий лексичний аналізатор. Він називається токенайзером або лексером. [3].

Токенайзер розбиває всі вхідні дані на токени - окремі символи, наприклад, слова. Отримані таким чином токени служать вхідними символами для парсера (див. рис. 1.1).

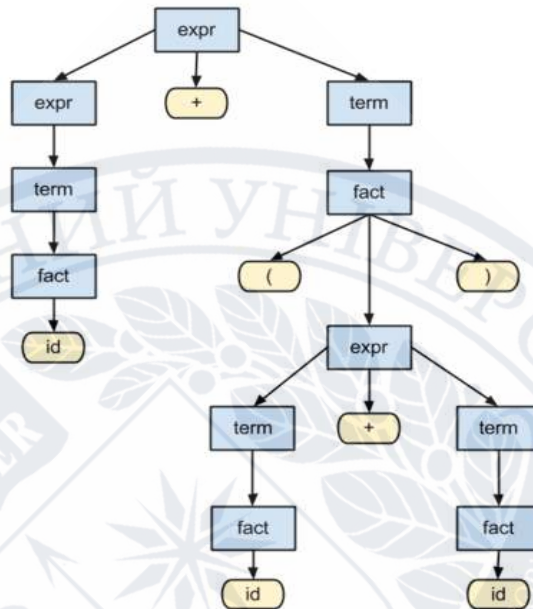


Рисунок 1.1 – Приклад дерева синтаксичного аналізу:
вираз – expr, термін – term, факт – fact, ідентифікатор – ID

Потім програма обробляє граматику вхідних даних, аналізує їх і створює синтаксичне древо. На цій основі йде подальша робота парсера з інформацією - генерація коду або вибірка за певними критеріями. [4].

Парсинг використовується для перетворення тексту в нову структуру в наступних випадках:

- Читання програмного коду - наприклад, Java, SQL та інших мов програмування. Цим займаються парсери баз даних. В даному випадку це працює так: парсер надає програмному компілятору структуру даних, за допомогою якої можна згенерувати машинний код.
- Читання HTML-коду за допомогою парсера вихідного коду сторінки. Для комп'ютера HTML-код - це рядки символів, які повинні бути проаналізовані парсером в браузері. Парсер надає опис веб-сторінки як структури даних, яка потім компонується в потрібному порядку і виводиться на екран комп'ютера.

- Аналіз XML-розмітки. Спеціальні парсери XML відповідають за аналіз цього типу документів і готують містити в них інформацію для подальшого використання.
- Читання URL-адрес і інтернет-протоколів HTTP і HTTPS. Парсери розбивають складні схеми URL-адрес і протоколів на ієрархічну структуру.
- Пошукові системи. За допомогою парсерів пошукової системи, боти вибирають релевантний для них текст з контенту сайту. Після первинної вибірки дані обробляються, а результати аналізу стають доступні для перегляду. [5].

У цифровому маркетингу парсинг застосовують, щоб зібрати і проаналізувати певну інформацію з контенту потрібних сайтів.

Парсинг сторінок, провідний збір інформації з веб-контенту сайтів, використовується в різних сферах - для продажів, маркетингу, фінансів, електронної комерції, збору інформації по конкурентах і так далі. Активно його застосовують в наступних областях.(див. рис. 1.2).

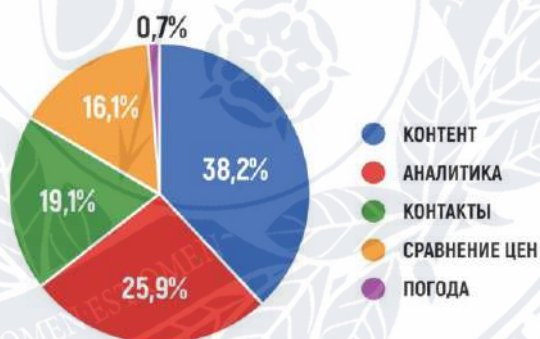


Рисунок 1.2 – Активне застосування по частоті, дослідження HubSpot роздрібного маркетингу

У роздрібній торгівлі існує безліч можливостей використання парсинга. Наприклад, моніторинг цін конкурентів або аналітика ринку, де парсинг використовують для обробки даних і вилучення з них цінної для маркетологів інформації.

Так, для електронної комерції може знадобитися незліченна кількість зображень і описів товарів. Їх не можна створити за пару-трійку днів, так як навіть скопіювати і вставити кожен займе певний час. Набагато простіше і швидше створити парсинг і швидко «вицепити» все потрібне. Або взяти аналітику ринкових цін - регулярний парсинг веб-сторінок конкурентів допоможе своєчасно помічати і враховувати всі зміни на ринку. [6].

Отже, парсинг - що це таке? Це процес аналізу вмісту сайту і вищлювання потрібної інформації. Використовуючи вищеписані відомості, можна наповнювати свої сайти великою кількістю контенту автоматично. А це дає можливість виграти час для створення великого масиву потрібної інформації та надати її в найкоротші терміни.

Розберемо парсинг на прикладі. Уявімо, що ми прийшли в бібліотеку і хочемо швидко переписати всі статті з великої радянської енциклопедії і розмістити їх у себе на сайті. І ось ми сидимо дні на проліт і вивчаємо кожен том, кожную статтю, шукаємо інформацію. А що якби одним натисканням кнопки хтось за нас зробив би всю цю роботу?

Або ми знайшли цікавий сайт з рецептами і хочемо швидко і повністю скопіювати їх на свій сайт. Це не проблема, коли сторінок всього кілька, але що якщо їх тисячі або навіть сотні тисяч? Як автоматизувати процес, що б не втратити час? [6].

Або інший приклад, більш наближений до реальності, що пов'язаний із освітнім процесом Донецького національного університету імені Василя Стуса.

У головному корпусі університету є структурний підрозділ як відділ кадрів. Працівники цього підрозділу постійно опрацьовують великий масив інформації, що потребує чимало зусиль з боку працівника і самого комп'ютера. Але за допомогою парсеру, у них була б чудова можливість його використати у роботі, при цьому скоротити час для пошуку тих самих лише потрібних інтернет-ресурсів та місце для зберігання інформації. Адже, коли проводиться синтаксичний аналіз за ключовими словами чи словосполученнями, інформація, яка уже оброблена автоматично зберігається або у хмарне сховище, або на

комп'ютер користувача, що не потребує постійного підтвердження чи саме в це місце потрібно її зберігати для подальшої обробки.

Це і є процес парсинга, тільки аналіз і збір інформації відбувається не з книг, а виключно з інтернет ресурсів. Процес парсинга виконується спеціальною програмою-парсером. Вона діє за певними правилами і алгоритмами, які закладає розробник. Наприклад, можна зібрати всі статті з вікіпедії або отримати телефони з сайту оголошень Avito і це лише мізерно мала частина можливостей парсингу. Об'єктом парсингу може бути блог, довідник, інтернет-магазин, форум та й будь-який сайт який можна побачити в інтернеті. Звичайно, є винятки, наприклад коли вся інформація зашита в Flash ролире. У будь-якому випадку перед парсингом потрібно вивчити ресурс-донор.

Результатом парсинга може бути текстовий файл, файл Ексель, csv, HTML файли, каталог з картинками або відео і будь-який інший формат на вимогу.

Виходить, що всі процеси, де застосовується синтаксичний аналіз, використовують парсинг. Це різні автоматизовані перекладачі з однієї мови на іншу, це і транслятори мов програмування, які формують програмний код на машинно-орієнтовану мову, це і мова SQL-запитів і тому подібні процеси.

Як неважко зрозуміти, парсинг в переважній більшості випадків за допомогою будь-якої програми, так як вручну або візуально це проробляти досить трудомістким. Ці програми називають парсером. [7].

Запити на послуги парсинга популярні на фріланс-біржах або форумах, присвячених тематиці SEO.

Дані, які найчастіше отримують за допомогою парсинга, можуть бути, наприклад, добірка статей певного автора, або підбір даних на певну тематику.

Також для аналізу часто парсять:

- Видачу пошукових систем;
- Пошукові запити, семантику;
- Інформацію по сайтам конкурентів (CMS, структуру, СЯ, ...);
- Дані з медіа-ресурсів, різних форумів;
- Даних з інтернет-магазинів - товарів, цін і т.п.

Систематизація і формування даних в форматі, звіту, таблиці або іншого зрозумілого документа.

Наведемо ще один приклад. Наприклад, є сайт-каталог промислових підприємств. Обробивши цей сайт за допомогою парсера, можна отримати дані у вигляді звіту або таблиці приблизно такого формату:

В SEO-просування парсинг найчастіше використовують для отримання контенту для подальшого рерайта або репостінга або для пошуку будь-яких веб-ресурсів - форумів, блогів, Email-адрес. Також популярний парсинг зовнішніх посилань, для аналізу сайтів-конкурентів і виявлення доступних трастових сайтів.

Парсинг (Parsing) - це прийняте в інформатиці визначення синтаксичного аналізу. Для цього створюється математична модель порівняння лексем з формальною граматиною, описана одним з мов програмування. Наприклад, PHP, Perl, Ruby, Python. [8].

Коли людина читає, то, з точки зору науки філології, він робить синтаксичний аналіз, порівнюючи побачені на папері слова (лексеми) з тими, що є в його словниковому запасі (формальної граматики).

Сфера застосування таких програм, як парсер, дуже широка, але всі вони працюють практично за одним алгоритмом.

На ділі зазвичай проводиться парсинг більш ніж одного ресурсу або сторінки сайту. Так що процес міг би затягнутися на досить значний час, якби алгоритм виконувався послідовно для кожної нової сторінки. При цьому ресурси комп'ютера, на якому працює парсер, використовувалися б лише в малому ступені. Рішення проблеми – багато-парсер, здатний обробляти величезну кількість сторінок і задіяти ресурси вашого комп'ютера максимально раціональним чином. Таким чином час парсинга скорочується в десятки разів.

Наприклад, багатопотоковий парсер від компанії AVADA MEDIA, що встановлюється на локальний комп'ютер або сервер – це рішення, більш підходяще для великих компаній.

Ще одне застосування парсинга для автобізнесу – парсер великих аукціонів США (див. рис. 1.3) [8].



Дуже потрібен парсер власникам інтернет-магазинів. З його допомогою вони отримують опис товарів, задаючи модель. В іншому випадку це довелося б щоразу робити вручну, що займало б сотні годин.

Компанія AVADA MEDIA пропонує рішення для інтернет-торгівлі – це може бути як хмарний парсер, так і завантажувати програмне забезпечення.

Окрема сфера застосування – парсинг валютних курсів і курсів криптовалют. Для тих, хто торгує на біржі Forex, важливо вчасно отримувати актуальну інформацію про зміни курсів. Допоможе спеціальна програма-біржі парсер-бот. [9].

Ця програма не лише збирає актуальну інформацію, а й торгує на біржі в рамках заданих користувачем змінних – біржовій комісії, бажаного прибутку.

Активно застосовуються боти і в області ставок на спортивні події, в тому числі кібердисципліни. В такому випадку програма-парсер займається збором коефіцієнтів в різних букмекерських конторах.

В даному випадку, користувач задає ареал пошуку – наприклад, певні дисципліни, змагання або дати. Бот шукає коефіцієнти і самостійно здійснює ставки в заданих користувачем прикладах. Робити це можна і вручну, однак займає набагато більше часу. Крім того, бот працює цілодобово, так що ви можете робити ставки на події в реальному часі, що відбуваються в інших часових поясах.

Компанія AVADA MEDIA розробляє готові рішення під ключ для сфери онлайн-беттінга і торгівлі валютами з урахуванням ваших вимог до програми. Їх фахівці також розробляють парсери під конкретні вузькі завдання для будь-якого бізнесу.

В даному випадку, парсинг даних з сайту – це процедура збору інформації по заданих параметрах. Програма-парсер сайтів проводить синтаксичний аналіз заданого онлайн-ресурсу і надає інформацію в заданому вигляді. [9]

Парсити сайт можна за допомогою самих різних парсерів – багатопоточних і універсальних, а також вузькоспеціалізованих, орієнтованих на конкретні завдання. Парсери сайтів за принципом роботи трохи схожі на пошукові боти, використовувані відомими пошуковими системами. Однак вони зазвичай парсять сайти по заданих параметрах і збирають сам контент з можливістю його використання, а не надають інформацію про нього.

Парсер контенту з сайту може зібрати його з будь-яких відкритих для людей і пошукових роботів джерел в мережі. Це можуть бути каталоги, інтернет-форуми, сайти оголошень, магазини, сайти-візитки, блоги, корпоративні портали, маркетплейси і багато іншого. Наприклад, власники інтернет-магазинів активно використовують парсери сайтів, щоб автоматизувати процедуру збору характеристик і фотографій розміщуються товарів на офіційних сайтах виробників і дистриб'юторів. Тобто, автоматизують роботу, яка зайняла б у людини в десятки разів більше часу і зажадала б значно більших коштів [10].

Наведемо конкретний приклад, пов'язаний з інтернет-магазином.

Припустимо інтернет-магазин включає в себе окремі картки товарів. Якщо заповнювати такі вручну, то на процес може піти чимало часу. Коли асортимент значний, то доведеться докласти багато зусиль. Вручну робити це нераціонально і довго.

Програми для парсинга оптимізують процес до мережі. З їх допомогою добитися бажаного можна буде всього за «пару кліків». Це дуже зручно, особливо алгоритм підходить тим інтернет-магазинах, які реалізують продукцію окремого бренду з офіційним сайтом, де зосереджені тисячі найменувань продукції. [10].

Використовувати програму парсинг при інтернет-магазині або в інших сферах діяльності дійсно дуже зручно. Адже вона дозволяє:

- Зібрати і швидко скопіювати інформацію з інших сайтів;
- Скопіювати дані і постійно підтримувати актуальність відомостей;
- Зібрати і обробити великі масиви інформації з подальшим розміщенням на особистому ресурсі.

Незалежно від того на якій формальній мові програмування написаний парсер, алгоритм його дії залишається однаковим:

- вихід в інтернет, отримання доступу до коду веб-ресурсу і його скачування;
- читання, вилучення та обробка даних;

-уявлення витягнутих даних в легкотравному вигляді - файли.txt, .sql, .xml, .html і інших форматах.

В інтернеті часто зустрічаються вирази, з яких випливає, ніби парсер (пошуковий робот, бот) подорожує по Всесвітній мережі. Але найчастіше ця програма ніколи не покидає комп'ютера, на якому вона інстальована.

Цим парсер докорінно відрізняється від комп'ютерного вірусу - автономної програми, здатної до розмноження, хоча по суті своєї роботи він схожий на трояна. Адже він отримує дані, іноді конфіденційного характеру, не питаючи бажання їх власника.

Збір інформації в інтернеті - трудомістка, рутинна, що забирає багато часу робота. Парсери, здатні протягом доби перебрати більшу частину веб-ресурсів в пошуках потрібної інформації, автоматизують її. [11].

Найбільш активно «парсять» всесвітню мережу роботи пошукових систем. Але інформація збирається парсером і в приватних інтересах. На її основі, наприклад, можна написати дисертацію. Парсинг використовують програми автоматичної перевірки унікальності текстової інформації, швидко порівнюючи вміст сотень веб-сторінок із запропонованим текстом.

Без програм парсинга власникам інтернет-магазинів, які потребують сотні однотипних описів товарів, технічних характеристик та іншої інформації, яка не є інтелектуальною власністю, було б важко вручну заповнювати характеристики товарів.

Можливістю «спарсити» чужий контент для наповнення свого сайту користуються багато веб-майстрів і адміністраторів сайтів. Це виправдано, якщо потрібно часто змінювати контент для подання поточних новин або інший, що швидко змінюється інформації. [12].

Парсинг – так звана «паличка-виручалочка» для організаторів спам-розсилок по електронній пошті або каналів мобільного зв'язку. Для цього їм треба запуснути «бота», який подорожуватиме та буде збирати «телефони, адреси, явки».

Ну і господарі деяких, особливо недавно організованих веб-ресурсів, люблять наповнити свій сайт чужим контентом. Правда, вони ризикують, оскільки пошукові системи швидко знаходять і банять любителів копіпаста.

Звичайно ж, парсери не читають текст, вони всього лише порівнюють запропонований набір слів з тим, що виявили в інтернеті і діють за заданою програмою. Те, як пошуковий робот повинен надійти зі знайденим контентом, написано в командному рядку, що містить набір букв, слів, виразів і знаків програмного синтаксису. Такий командний рядок називається «регулярний вираз». Російські програмісти використовують жаргонні слова «маска» і «шаблон».

Щоб парсер розумів регулярні вирази, він повинен бути написаний на мові, що підтримує їх в роботі з рядками. Така можливість є в PHP, Perl. Регулярні вирази описуються синтаксисом Unix, який хоча і вважається застарілим, але широко застосовується завдяки властивості зворотної сумісності.

Синтаксис Unix дозволяє регулювати активність парсинга, роблячи його «ледачим», «жадібним» і навіть «зверхскупим». Від цього параметра залежить довжина рядка, яку парсер копіює з веб-ресурсу. Зверхскупий парсинг отримує весь контент сторінки, її HTML-код і зовнішню таблицю CSS. [13].

1.2 Етапи розвитку програм-парсерів в Україні

В даний час всі процеси, де застосовується синтаксичний аналіз, використовують парсери - програми для проведення візуального або програмно-автоматизованого синтаксичного і лексичного аналізу або розбору будь-якого документа з метою вилучення з нього необхідних даних. Це і різні автоматизовані перекладачі з однієї мови на іншу, і транслятори мов програмування, які формують програмний код на машинно-орієнтовану мову, це і мова SQL-запитів і тому подібні застосування.

Для створення парсерів використовуються найрізноманітніші мови програмування. [14].

Створення програми-парсера не вимагає серйозних знань про мову програмування. Необов'язкові і фундаментальні відомості про технології. Але щось знати все-таки необхідно. Отже, щоб правильно створити парсинг, потрібно засвоїти етапи розвитку та формування програм-парсерів в Україні та світі.

У даному випадку, в хронологічному порядку перерахуємо найважливіші етапи створення та розвитку парсерів як інструменту аналізу даних:

- **1960.** Випущена специфікація ALGOL 60, в якій вперше описано мову з блочною структурою. Комітету ALGOL добре відомо, що ніхто не знає, як парсити таку мову. Але вони вважають, що, якщо вони детально описали б мову з блочною структурою, то для нього буде винайдений аналізатор - парсер. Це був ризикований підхід, який згодом окупився;

- **1961.** Нед Айронс випускає свій парсер ALGOL. Фактично, алгоритм Айронса є першим подібним аналізатором, який був описаний у пресі. Парсер Неда є ліво-рекурсивним (форма рекурсивного спуску) парсером. На відміну від сучасного рекурсивного спуску, алгоритм Айронса носить загальний характер і є синтаксично-керованим. «Загальний» означає, що вони можуть розібрати, що написано в БНФ (форма Бекуса - Наура - формальна система опису синтаксису, в якій одні синтаксичні категорії послідовно визначаються через інші категорії). «Синтаксично-керований» (декларативний) означає, що аналізатор фактично створюється з БНФ - парсер не потрібно створювати самому;

- **1961.** Майже одночасно, з'являються самокодовані (тобто з можливістю внесення змін до початкового коду) підходи до реалізації ліворекурсивних алгоритмів. Зараз їх розглядають як рекурсивний спуск. Протягом наступних років самокодовані підходи стають все більш популярними для ліворекурсивних аналізаторів, ніж синтаксично-керовані алгоритми. Важливу роль відіграли такі три фактори: [14].

У **1960-х** пам'ять і CPU дуже обмежені. Ручне кодування (hand-coding) окупається навіть тоді, коли виграш від його застосування малий.

Чистий ліворекурсивний аналіз - дуже слабкий метод синтаксичного аналізу. Ручному кодуванню часто доводиться долати ці обмеження. Це твердження вірне як в 1961, так і в наші дні.

Ліворекурсивний аналіз добре працює в поєднанні з ручним кодуванням - вони доповнюють один одного.

- **1965.** Дональд Кнут (Don Knuth) винаходить алгоритм LR. Вчений в першу чергу зацікавлений у математичній стороні завдання. Кнут описує алгоритм аналізу, але його підхід вважається непрактичним;

- **1968.** Джей Ерлі (Jay Earley) винаходить алгоритм, названий на його честь. Як і алгоритм Irons, алгоритм Ерлі є синтаксично-керованим і носить загальний характер. На відміну від алгоритму Irons, не використовує метод пошуку з поверненням. Основна ідея Ерлі полягає в тому, щоб відстежувати етапи роботи алгоритму в таблицях. Алгоритм Ерлі привабливий, але має три основних недоліки: по-перше, є помилка в обробці правил нульової довжини; по-друге, при право-сторонньої рекурсії необхідно застосовувати алгоритм двічі; по-третє, щоб створити таблиці, необхідно вести облік системних ресурсів, що, за мірками апаратних засобів, є досить складним завданням; [15].

- **1969.** Френк ДеРемер (Frank DeRemer) описує новий варіант LR Кнута. Для алгоритму LALR ДеРемера необхідний тільки стек і таблиця станів, розмір якої можна легко змінити;

- **1972.** Ахо (Aho) і Алмані (Ullmann) описали простий спосіб виправити помилку правила нульової довжини в оригінальному алгоритмі Ерлі. На жаль, це виправлення вимагає навіть більше системних ресурсів, ніж алгоритм Ерлі;

- **1975.** Белл Лебс (Bell Labs) перетворює свій компілятор мови C з самокодуючого рекурсивного спуску в алгоритм LALR ДеРемера;

- **1977.** Виходить перша «Книга дракона». Цей, який невдовзі став класичним, адже підручник названий так через обкладинку, на якій зображений лицар, що кидає виклик дракону. На спис лицаря вигравірувані літери «LALR».

Надалі, говорити легковажно LALR - значить заплямувати герб теорії синтаксичного аналізу;

- **1979.** Bell Labs випустила нову версію UNIX - сьому. V7 включає в себе безумовно найбільш повний, корисний і доступний інструментарій для розробки компіляторів. Центральне місце займає інструментарій Yacc, генератор синтаксичних аналізаторів на основі LALR. З невеликою працею Yacc парсить свою власну мову введення, а також мову основного компілятора V7 - переноситься мова компілятора C. Здається, що після двох десятиліть досліджень, проблема синтаксичного аналізу нарешті вирішується; [15].

- **1987.** Ларрі Уолл (Larry Wall) являє Perl 1. Perl дозволяє вирішувати більш складні завдання, чим відрізняється від уже існуючих мов. Ларрі активно використовує LALR - наскільки відомо автору, частіше, ніж будь-хто до або після нього;

- **1991.** Джуп Лео (Joop Leo) виявляє спосіб прискорити справа наліво рекурсії в алгоритмі Ерлі. Алгоритм Лео є лінійним практично для будь-якої: як однозначною, так і неоднозначною граматики, що представляє практичний інтерес. Апаратне забезпечення в 1991 на шість порядків швидше, ніж в 1968 році, так що питання обліку системних ресурсів став не настільки важливий. Однак, якщо справа стосується швидкості, то виграє алгоритм Ерлі. Алгоритм Лео виявився важливим відкриттям, але його практична реалізація з'явиться лише через 20 років;

- **1990-і.** Алгоритм Ерлі забутий. Всі прихильники LALR задоволені, але не всі. Користувачі LALR роблять неприємні відкриття. У той час як LALR автоматично генерує свої аналізатори, їх налагодження настільки важке, що простіше написати парсер самому. Після налагодження їх LALR аналізатори при правильних вхідних даних працюють швидко. Але майже всі, що вони говорять користувачам про некоректні вхідні дані - лише повідомлення про неправильний формат без вказівки додаткової інформації. За словами Ларрі, LALR «швидкий, але дурний»;

- **2000.** Ларрі Уолл приймає рішення про радикальне перевизначенні Perl - виданні Perl 6. Він навіть не розглядає питання про повторне використання LALR;

- **2002.** Ейкок (Ayscock) і Хоспул (Horspool) опублікували результати своїх дослідів зі створення швидкого, практичного парсеру Ерлі. Однак при цьому не використовувалося поліпшення, запропоноване Джуа Лео - вони, здається, не знають про нього. Їх власний метод прискорення обмежений в своїх показниках і складності, які він несе з собою, вони можуть бути навіть контрпродуктивними при оцінці часу. Але в цій роботі цінних виявилось рішення помилки правила нульової довжини. І на цей раз вона не вимагає ніякого додаткового обліку системних ресурсів;

- **2006.** GNU оголошує, що був переписаний парсер компілятора GCC. Протягом трьох десятиліть, компілятори мови C промислового флагмана використовували LALR як парсер - доказ твердження, що LALR і серйозний алгоритм еквівалентні. Тепер GNU замінює LALR технологією, яку той замінив чверть століття тому: рекурсивним спуском;

- **З 2000 і до сьогоднішнього дня.** З відступом від LALR приходить крах престижу теорії синтаксичного аналізу. Через півтора століття, ми прийшли до того, з чого почали. Якщо взяти оригінальний алгоритм Ned Irons 1961 року, що змінить імена і дати і перекладемо код з суміші асемблера і ALGOL в Haskell, то ми б легко змогли випустити його в наші дні і представити як новий і революційний підхід.

- **Парсер Marpa.** Оригінальне, давно занедбане уявлення алгоритму Ерлі - ефективний, практичний, загальний і синтаксично-керований парсер - тепер, по суті, цілком можливо застосовувати на практиці. [16].

Ейкок і Хоспул вирішили проблему правила нульової довжини. Джуа Лео розробив прискорення для правобічної рекурсії. І питання обліку системних ресурсів розв'язався сам по собі: машинні операції в даний час в мільярд разів швидше, ніж в 1968 році, і швидкість, ймовірно, вже не така велика - в даний час головною проблемою стали кеш-промахи.

Але в той же час як вирішилися проблеми оригінального алгоритму Ерлі, з'явився новий метод. З таким же потужним, як алгоритм Ерлі, синтаксично-керований підхід буде працювати набагато ефективніше, ніж з лівостороннім парсером. В даний час не так багато програмістів використовують чисто декларативний парсер через сумного досвіду роботи з LALR. Як то кажуть, раз обпечешся, іншим разом остережешся. [17].

Отже, щоб правильно працювати з програмою-аналізатором, потрібно засвоїти наступне:

- *для початкового алгоритму функціонування програми потрібен ретельний аналіз вихідного коду веб-сторінки, що є донором. Тут не обійтися хоча б без середніх знань технологій верстки. Це і мова JavaScript.*
- *щоб зануритися в тему глибше, потрібно вивчити технологію під назвою DOM. Вона дає можливість дуже ефективно працювати з ієрархією веб-сторінки.*
- *найважчий етап - написання парсеру. Тут потрібно володіти інструментом для обробки тексту. Досвідчені програмісти найчастіше використовують для цієї мети регулярні вирази, які є досить потужним засобом. Але це під силу далеко не кожному розробнику. Тут потрібна особлива мислення. Оптимальним рішенням буде використання вже готових бібліотек, які створювалися спеціально під парсинг. Це упакований програмний код, який вже містить всі функції для аналізу.*
- *дуже бажано розбиратися в об'єктно-орієнтованому програмуванні, яке підтримується будь-якою мовою програмування.*
- *завершальний етап обробки результатів аналізу передбачає, що дані будуть структуровані і збережені. Тут не обійтися без знань баз даних.*
- *потрібні знання і володіння функціями, службовцями для роботи з файлами. [18].* Адаже дані потрібно буде записувати в ці самі файли, а потім, можливо, конвертувати в формат електронних таблиць. Після успішної обробки даних, їх необхідно зберегти в тому вигляді, який можна буде в подальшому обробляти. Так як сайт складається не з однієї сторінки, а з безлічі, алгоритм

повинен вміти переходити на наступні сторінки. Парсер надає інформацію в певному виді, який так само задається розробником програми.

Процес парсинга в загальному випадку можна розділити на три етапи:

- Аналіз вихідних даних, відбір документів для подальшого опрацювання.
- Розбір відібраних документів на складові частини
- Вибір необхідної інформації і збереження даних в необхідному форматі. [19].

Весь процес інтернет-парсинга в широкому випадку можна розділити на кілька етапів:

- Отримання коду інтернет сторінки
- Аналіз отриманих даних
- Обробка і формування результату
- Висновок результату в файл або на екран

Парсинг html-сторінки вдає із себе процес, який можна розбити на три етапи:

1. *Отримання початкового коду веб-сторінки* - скачати програмний код тієї сторінки сайту, з якої необхідно витягти інформацію. У різних мовах для цього передбачені різні способи, наприклад, в PHP 18 найчастіше використовують бібліотеку cURL або ж вбудовану функцію `file_get_contents`.

2. *Витяг з html-коду необхідних даних*. Отримавши сторінку, необхідно обробити її - відокремити звичайний текст від гіпертекстової розмітки, вистроїти ієрархічне дерево елементів документа, коректно зреагувати на невалідний код, виокремити зі сторінки саме ту інформацію, заради якої і відбувається весь цей процес. Можна, звичайно ж, використовувати для цього регулярні вирази, проте є більш зручний шлях - спеціалізовані бібліотеки.

3. *Фіксація результату*. Благополучно обробивши дані на сторінці, потрібно їх зберегти в необхідному вигляді для подальшої обробки.

Спарсені дані зазвичай заносяться в базу даних, однак є й інші варіанти. Іноді потрібно записати в CSV-файл або будувати ієрархічні JSON-структури,

іноді конвертувати в excel-таблицю, а може навіть згенерувати динамічний rss-потік. Як правило, потрібно спарсити не одну сторінку сайту-донора, а безліч, може навіть і все. У цьому випадку після проходження кроків 1-3 у алгоритм парсера повинен бути закладений перехід на наступну сторінку сайту, щоб і з неї вилучити весь необхідний матеріал. [20].

Обхід всіх потрібних сторінок сайту забезпечується різними способами:

- По-перше, обробляючи чергову сторінку, парсер можна навчити не тільки отримувати необхідні дані, але і заносити в свою базу даних всі внутрішні посилання, що зустрічаються на шляху. Звертаючись до свого сховища лінків, програма послідовно відвідує сторінки сайту, до тих пір поки не обійде їх всіх;
- По-друге, при первинному аналізі сайту найчастіше можна простежити логіку формування url для сторінок. І потім, генерувати адреси відповідно до виявлених закономірностей;
- По-третє, деякі парсери розраховані, як не дивно, на «ручний» обхід веб-ресурсу. Користувач, клікаючи по посиланнях, сам вирішує які 19 сторінки відвідувати, які ні. А програма у фоновому режимі запам'ятовує необхідні дані. Зрозуміло, поєднувати різні методи ніщо не заважає.

Але перед тим як парсити необхідні дані, потрібно знати чи визначена його правомірність. [21].

Законодавчого визначення поняття парсингу немає, так само як немає його і для суміжних понять краулінг (обхід обмежень сайту) і скрайпінг (об'єднані в одну програму краулінг і парсинг).

Оскільки парсинг – це один із способів збору інформації, то цей процес підпорядковується правилам, встановленим для поводження з інформацією, зокрема, Закону України «Про інформацію».

Цей закон ділить інформацію за двома критеріями: за змістом і за режимом доступу. Причому режим доступу визначається, в тому числі, за змістом.

Будь-яка інформація є відкритою, крім тієї, що віднесена до інформації з обмеженим доступом. Відповідно, можна використовувати будь-яку інформацію, яка не обмежена в доступі.

У свою чергу інформація з обмеженим доступом поділяється на конфіденційну, таємну та службову. Визначення кожного виду містяться в Законі України «Про доступ до публічної інформації». Нас цікавить в першу чергу визначення конфіденційної інформації, тому що, найімовірніше, суперечки про правомірність парсингу будуть зводитися до того, чи була «зпарсена інформація» конфіденційною. До того ж таємна або службова інформація точно не буде перебувати у відкритому доступі на сайті. [22].

Щоб розібратися з такою ситуацією, наведемо приклади та кейси, що торкаються цього питання:

Збір баз даних користувачів. Існує безліч сервісів, які збирають бази даних користувачів. Наприклад, збір передплатників певної спільноти в соцмережі. Почнемо з того, що майже вся інформація про фізичну особу є конфіденційною і може бути використана тільки за її згодою. Користувач надає таку згоду, коли реєструється на сайті. Сайт стає розпорядником персональної інформації. Такою інформацією є відомості чи сукупність відомостей, за якими особа може бути ідентифікована. Частковий виняток становить інформація про держслужбовців та інших публічних осіб (музикантів, акторів, спортсменів).

Розпорядник персональної інформації також повинен надати згоду на використання персональних даних.

Якщо коротко, щоб парсинг був законним, парсити потрібно або деперсоніфіковані дані, або отримувати згоду розпорядника інформації. (Наприклад, у 2010 році хакер Ендрю Ауернхаймер знайшов пролом в системі безпеки на веб-сайті AT&T і витягнув адреси електронної пошти користувачів, які відвідували сайт зі своїх iPad. Скориставшись недоліком системи безпеки і парсингом, Ауернхаймер зміг отримати доступ до тисяч електронних адрес з сайту AT&T. Ауернхаймер був визнаний винним у несанкціонованому доступі до сервера AT&T і присвоєнні чужих даних.) [23].

Якщо мова йде про інформацію, яка не є персональною, вона може вважатися конфіденційною, тільки якщо її власник визначив її як таку. Зазвичай сайти пишуть політики конфіденційності, з якими користувач ознайомлюється і

які погоджується дотримуватися при реєстрації. Інформація, доступ до якої можна отримати тільки пройшовши процес авторизації, майже напевно буде конфіденційною і не може бути використана без згоди власника.

Але навіть якщо інформація доступна без авторизації, це не означає, що інформація знаходиться у вільному доступі. Перед парсингом слід перевірити, чи не заборонено Правилами використання сайту (якщо такі є) використання даних сайту. (Наприклад, *Ryanair* подали в суд на *PR Aviation* за те, що вони збирали інформацію про ціни на авіаквитки, що було заборонено Правилами користування *Ryanair*. Сайт *Ryanair* вимагає ознайомитися і погодитися з Правилами користування шляхом постановки галочки у відповідному полі перед пошуком авіаквитків. Особа автоматично зобов'язується дотримуватися правил користування сайтом при пошуку. Одне з правил користування говорить, що «забороняється використовувати цей сайт ... бази даних, функції або вміст з будь-якою метою, крім особистої некомерційної мети». Відповідно, *PR Aviation* порушили правила користування.) [24].

Більш того, сайт може висловити незгоду на використання його інформації й іншими способами. Наприклад, шляхом блокування IP адрес парсеру або ж шифрування даних. Обхід подібних блокувань буде незаконним, навіть якщо вони були встановлені вже після початку парсингу. (Наприклад, *Craigslist* – сервіс для публікації різних оголошень. *3Taps* парсив дані з сайту *Craigslist* для створення аналітики *big data*. *Craigslist* через деякий час заблокував IP адреси *3Taps*, але останні використовували проксі для продовження доступу до сайту. *Craigslist* звернулися до суду з приводу порушення їх прав і виграли справу. Ключовим моментом став саме факт обходу блокувань. *3Taps* виплатили *Craigslist* 1 мільйон доларів.)

Авторське право. Окремо варто згадати парсинг інформації, яка містить дані, що охороняються авторським правом. Таку інформацію парсити потрібно вкрай обачно, оскільки використання такої інформації в своєму проєкті, найімовірніше (залежить від виду ліцензії), буде порушенням авторських прав. (Наприклад, *Associated Press* подали в суд на *Meltwater* (розробник програми для

збору новин) за парсинг статей, частина з яких були захищені авторським правом. Хоча факти, тобто новини, не підлягають охороні авторським правом, суд вирішив, що копіювати авторський виклад статей незаконно.)

Порушення роботи сайту. Навіть за умови правомірності парсингу, його здійснення не повинно підривати нормальну роботу сайту, який парсять. (Наприклад, *QVC* (телевізійний рітейлер) подали в суд на *Resultly* (додаток-магазин) через те, що пошукові боти *Resultly* перевантажили сервери *QVC* з відключенням електроенергії, що призвело до збитків у 2 мільйони доларів. Суд виправдав *Resultly*, на тій підставі, що вони не мали наміру завдати шкоди. Проте, в схожих ситуаціях ризик понести відповідальність за порушення роботи сайту все ж залишається.) [25].

Після таких збитків із парсерами виникає невід’ємне питання, як легально парсити і як захиститися від цього збору інформації.

Наприклад, повністю попередити судові тяжби через парсинг неможливо, оскільки ця сфера не врегульована. Проте, можливо звести їх ймовірність до мінімуму.

Досить дотримуватися цих простих правил:

1. Не парсити інформацію з обмеженим доступом або захищену авторським правом.
2. Перевірити, чи не заборонений парсинг правилами користування сайту.
3. Якщо є можливість отримати згоду на парсинг, краще це зробити. До речі, такою згодою може бути API сайту.
4. Якщо після початку парсингу сайт вжив заходів до обмеження або зупинення парсингу, їх не слід ігнорувати або обходити.
5. Не порушувати роботу сайту.

Якщо дотриматися всіх цих правил, наші позиції, навіть в тому суді будуть досить переконливі. Якщо ж наш сайт став об’єктом парсингу, а ми того не бажаємо, потрібно здійснити наступні дії:

1. Заблокувати відповідні IP адреси.

2. Направити автору парсеру (якщо ми можемо встановити його) вимогу припинити парсинг.

3. Звертайтесь до суду з вимогою припинити порушення прав.

Але саме головне це - рекомендується прописати детальні Правила користування і Політики конфіденційності для нашого сайту. [26]

1.3 Мета, завдання та передумови впровадження програм-парсерів в закладах вищої освіти

Існує два основні методи парсинга: спадний і висхідний. Зазвичай вони розрізняються порядком, в якому створюються вузли синтаксичного дерева.

- Зверху-вниз: при низхідному методі парсер перевіряє наявність зверху - з початкового символу в коді і шукає відповідні йому синтаксичні зв'язки. Таким чином, дерево синтаксичного аналізу розростається зверху вниз, у напрямку більш детальної розбивки. [27].

- Знизу-вгору: висхідний парсер починає знизу, з самого нижнього ланцюжка довжиною, а потім встановлює все більші синтаксичні зв'язки. Це робиться до тих пір, поки не буде досягнуто початковий символ коду.

Але метод роботи - не найважливіше. Добре зроблений парсер - висхідний або спадний - буде розрізняти, яка інформація, наприклад, в рядку HTML необхідна. І відповідно до заздалегідь написаним кодом і правилами синтаксичного аналізу вибере потрібні дані і перетворює їх, наприклад, в формат JSON, CSV або навіть в таблицю Excel. [28].

Важливо відзначити, що сам парсер інформації не прив'язаний до певного формату даних. Це інструмент, який перетворює один формат в інший. А ось як він перетворює його і під що, залежить від поточних завдань.

Головна мета парсингу – перетворення кодової інформації з веб-сторінки в текстовий або інший вид, в якому вона може сприйматися більш зручно. Якщо спробувати пояснити, що це таке, простими словами, визначення буде приблизно таким - це ПЗ, що виділяє певні частини інформації з масиву даних. [29].

Алгоритм роботи парсера може відрізнятися в різних реалізаціях, але основний принцип залишається незмінним:

- Програма сканує дані, що надходять на вхід, будь то текст, веб-сторінка або інший набір інформації, і виокремлює з них деякі елементи.
- Що саме буде виділяти парсер з масиву даних - залежить від конкретного завдання. Зазвичай програми можна налаштовувати таким чином, щоб отримувати потрібні результати. [30].
- Правила пошуку найчастіше задаються регулярними виразами - рядками, складеними за певними правилами і дають програмі пояснення, що і як шукати.
- На основі зібраної інформації формується звіт або таблиця, в якій відображені всі отримані результати.

Головне завдання парсерів – автоматизація збору та систематизія даних. Подібні програми задіяні не тільки в рекламі. Вони використовуються в широкому ряді областей: від програмування до аналізу даних. Це означає, що парсер - інструмент в принципі багатофункціональний, і область його застосування не обмежена вузькими рамками.

До найбільш яскравим варіантів використання парсеру сайтів можна віднести:

- збір інформації для наповнення інтернет-магазинів: в такому випадку програма збирає неунікальні відомості, не захищені авторським правом, і використовує їх для складання карток товарів (наприклад, такими відомостями можна вважати зображення позицій);
- отримання різноманітного контенту з інших ресурсів для подальшого аналізу або використання в своїх цілях;
- моніторинг новин або оголошень: таке часто практикують люди, в роботі яких важливо оперативне отримання інформації: ріелтори, дилери старих автомобілів і т. д. [31].

На практиці компанії застосовують парсер в основному за кількома напрямками, і основні приклади його використання можна розбити на три групи:

Новини. Дуже часто парсер - знаряддя новинних компаній та інших організацій, яким важлива швидкість отримання даних. Програма, працюючи з величезною швидкістю, виокремлює з сотень сторінок потрібну інформацію і в результаті поставляє її співробітникам набагато швидше, ніж це зробив би людина.

Яскраві приклади фірм, які використовують парсер таким чином, - агрегатори новин, ріелторські агентства, компанії, що займаються перепродажем. [32].

Соціальні мережі. При веденні популярної групи в соцмережі або розкрутці через неї без парсеру складно обійтися. Його використовують для пошуку відомостей, які дозволять просунути свій аккаунт або сторонній ресурс, перенесення даних про користувачів і відбору потенційних клієнтів. [33]. Парсер дозволить проаналізувати, наприклад, склад учасників в співтоваристві або зібрати контактні відомості з тієї чи іншої групи користувачів. (див. рис. 1.4).

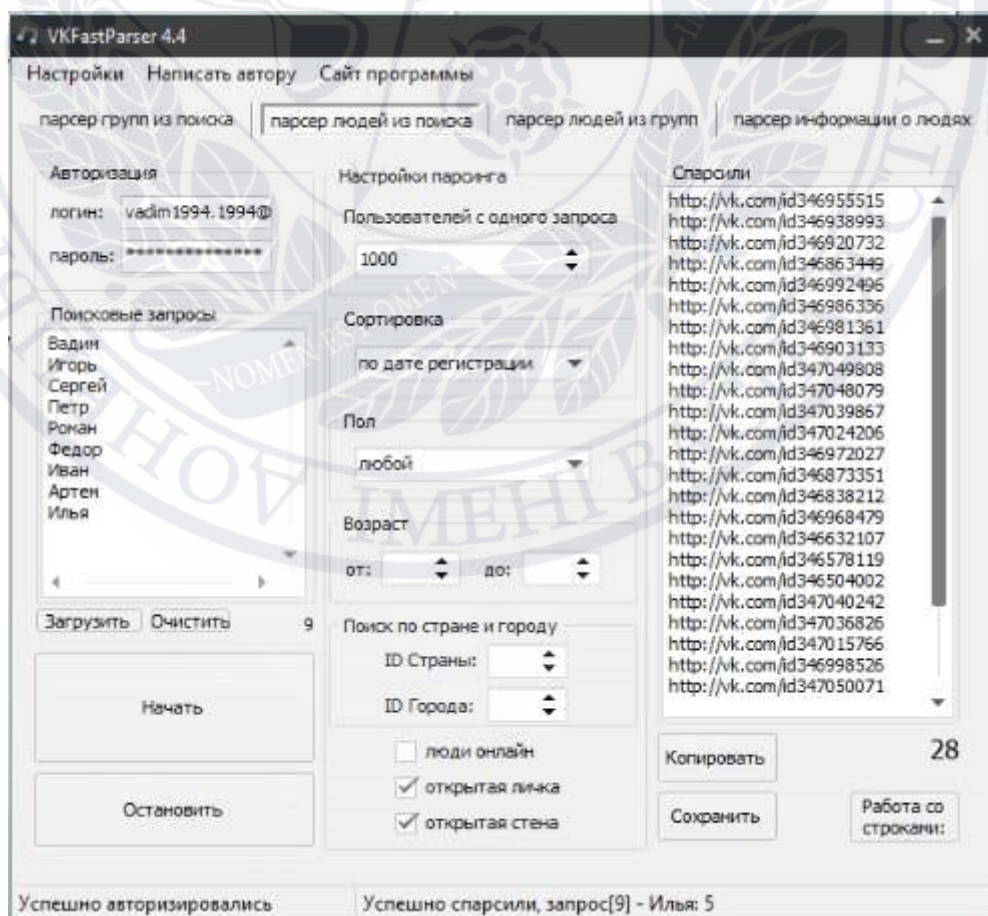


Рисунок 1.4 – Аналіз складу учасників в співтоваристві за допомогою парсера

Оновлення інформації. Таким чином парсери використовують з досить великою періодичністю: на відміну від новин, які можуть з'являтися з величезною швидкістю, важливі дані оновлюються не дуже швидко. Наприклад, туристичні компанії можуть запускати парсер, щоб дізнатися про зміну в умовах при відпочинку в будь-якому місці, а інтернет-магазини - щоб зрозуміти, чи не з'явився новий товар. [34].

Узагальнивши все, буде зрозуміло, що ДонНУ імені Василя Стуса потрібні програми-парсери, адже передумовою впровадження програм-парсерів в закладах вищої освіти є обробка великого масиву інформації в найкоротші терміни; у звіті не буде помилок; при синтаксичному аналізі, щоб було можна виставляти десятки параметрів, що потрібні при роботі з програмою, у результаті чого буде те, що конкретно потрібно; потрібна програма, де можна налаштувати так, щоб збирати дані, наприклад лише по понеділках; ці програми не тільки будуть збирати дані, а й будуть рекомендувати, що потрібно видалити, та які помилки виправити тощо. [35].

Висновки до розділу 1

Отже, парсер – це програма, що забезпечує парсинг, тобто синтаксичний аналіз контенту в мережі до певної математичної моделі, створеної на одній з мов програмування, зазвичай на Python, PHP або Java.

Алгоритм роботи програми-парсера такий: 1) програма отримує доступ до ресурсу в мережі інтернет; 2) відбувається завантаження коду сторінки; 3) читання і обробка інформації; 4) надання результату в зручних форматах – .html, .xml, .sql та інших.

Процес парсинга поділяється на три етапи: 1) аналіз вихідних даних; 2) розбір відібраних документів на складові частини; 3) вибір необхідної інформації і збереження даних в необхідному форматі.

Оскільки парсинг – це один із способів збору інформації, то цей процес підпорядковується правилам, встановленим для поводження з інформацією, зокрема, Закону України «Про інформацію».

Для того, щоб правильно спарсити інформацію, потрібно дотримуватися таких простих правил, при яких, не будуть порушуватися закони та права: 1) не парсити інформацію з обмеженим доступом; 2) перевірити, чи не заборонений парсинг правилами користування сайту; 3) якщо є можливість отримати згоду на парсинг, краще це зробити; 4) якщо після початку парсингу сайт вжив заходів до обмеження або зупинення парсингу, їх не слід ігнорувати або обходити; 5) не порушувати роботу сайту.

Головна мета парсингу- перетворення кодової інформації з веб-сторінки в текстовий або інший вид, в якому вона може сприйматися більш зручно.

Головне завдання парсерів – автоматизація збору та систематизія даних.

Основні приклади використання програми можна розбити на три групи: новини, соціальні мережі, оновлення інформації.

І в даному випадку, передумовою впровадження програм-парсерів в закладах вищої освіти є обробка великого масиву інформації в найкоротші терміни; у звіті не буде помилок; при синтаксичному аналізі, можна виставляти десятки параметрів, що потрібні при роботі з програмою, у результаті чого буде те, що конкретно потрібно; програму можна налаштувати так, щоб збирати дані, наприклад лише по понеділках; ці програми не тільки будуть збирати дані, а й будуть рекомендувати, що потрібно видалити, та які помилки виправити тощо. [35].

РОЗДІЛ 2

АНАЛІЗ ВИКОРИСТАННЯ ПРОГРАМ-ПАРСЕРІВ В ЗАКЛАДАХ ВИЩОЇ ОСВІТИ (НА ПРИКЛАДІ ДОНЕЦЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ ІМЕНІ ВАСИЛЯ СТУСА)

2.1 Аналіз комунікаційного процесу закладу вищої освіти

Сучасні мультимедійні комп'ютерні програми та інформаційно-комунікаційні технології відкривають доступ до нетрадиційних джерел інформації - електронних підручників, освітніх Web-сайтів. Це дає можливість підвищити ефективність розвитку пізнавальної самостійної діяльності й дати нові можливості для робочого зростання працівників.

Питанням інформатизації сучасного комунікаційного процесу й основам використання інформаційних технологій присвячена значна кількість досліджень. Під інформаційними технологіями розуміється широкий й не однозначний доступ для використання у комунікаційному процесі: баз даних, електронних таблиць, інформаційних мереж, які зводиться частіше до опису засобів роботи. [36].

Інформаційно-комунікаційні технології можна застосовувати як:

- засоби навчання: засоби, що вдосконалюють робочий процес;
- інструмент пізнання навколишньої дійсності і самопізнання;
- засоби комунікації;
- засоби автоматизації процесу обробки результатів експерименту і управління;
- засіб автоматизації процесів контролю і коригування результатів навчальної діяльності, тестування і психодіагностики;
- засоби організації інтелектуального дозвілля.

Інформаційні технології розвивають ідеї програмованого навчання, орієнтовані на локальні комп'ютери. В процесі цього використовуються готові навчальні і демонстраційні комп'ютерні проектні середовища.

Використання інформаційно-комунікаційних технологій у комунікаційному процесі може забезпечити передачу знань і доступ до різноманітної інформації нарівні, а іноді й інтенсивніше та ефективніше, ніж за традиційного пошуку. У разі впровадження програм-парсерів в освітній процес значно спрощуються такі значні за обсягом роботи організаційного характеру, як розробка і коригування планів, повсякденна і достовірна інформація про працівників тощо. Безумовно, подібна комп'ютерна мережа повинна бути інтегрована як методично, так і технологічно. Одним із напрямів застосування інформаційно-комунікаційних технологій в освіті є парсери в процесі якого відбувається одержання конкретної інформації за допомогою спеціалізованого середовища, заснованого на використанні інформаційних технологій, які забезпечують надання інформації.[36].

Інформаційно-комунікаційні технології у відділі кадрів дають можливість до залучення кожного працівника в активний пізнавальний процес, причому процес не пасивного оволодіння знаннями, а активної пізнавальної діяльності. Це можливість працювати спільно, в співпраці, в процесі розв'язування різноманітних проблем, проявляючи під час цього певні комунікативні уміння, можливість широкого спілкування під час роботи ц цими програмами.

Багато дослідників відзначають такі особливості інформаційно-комунікаційних технологій, як багатофункціональність, оперативність, продуктивність, насиченість, можливість швидкої й ефективної творчої само актуалізації працівників. Це не лише потужний засіб, що дозволяє в роботі працювати з інформацією, а з іншого боку, комп'ютерні телекомунікації - це особливе середовище спілкування один з одним, середовище інтерактивної взаємодії представників різних національних, вікових, професійних й інших груп користувачів незалежно від їхнього місця знаходження. Відрізняючись високим ступенем інтерактивності, комп'ютерні телекомунікації створюють унікальне

навчально-пізнавальне середовище, тобто середовище, що використовується для розв'язування різних дидактичних завдань.

І.Б. Горбунова відзначає, що нині увага фахівців зосереджена на розробці різних навчальних комп'ютерних моделей, моделюючих середовищ й різних видів програм для обчислювального експерименту.

Використання інформаційно-комунікаційних технологій у комунікаційному процесі впливає на методичну систему роботи на всіх її рівнях [36]:

- на рівні цілей роботи;
- на рівні змісту роботи;
- на рівні методів роботи.

Інформаційно-комунікативні технології дозволяють будувати процес комунікацій таким чином, що:

- у зміст роботи з парсерами включається вивчення стратегій пошуку інформації, в тому числі творчих; забезпечується аналіз і засвоєння діяльності, що включає пошук та переробку інформації;
- зміст робочого процесу будується з урахуванням реальних виробничих процесів.

Отже, інформаційно-комунікаційні технології дозволяють зробити комунікаційний та робочий процес більш інтенсивним, ефективним за рахунок реалізації можливостей мультимедіа комп'ютерних програм до дієвого й наочного подання інформації; створювати сприятливу атмосферу для спілкування, обміну й взаємодії як окремих працівників, так і освітніх систем.

Сучасна наука має різні погляди на природу, організацію, структуру комунікативної діяльності. Проте до її елементної бази відносять:

Суб'єкт комунікації – комунікант; предмет комунікації – комунікат; комунікативні засоби – знакові системи або засоби комунікації; комунікативний процес – процес установалення й підтримування контактів.

Сам процес комунікації формують фактори, які задаючи його форму та зміст, можна визначити наступним ланцюгом параметрів (див. рис. 2.1).



Рисунок 2.1 – Фактори, що формують процес комунікації

В сьогоденних умовах діяльності, комунікацію необхідно розглядати як підсистему вищого навчального закладу з її основними функціями:

- поширення стратегії розвитку ВНЗ;
- інтеграція й регулювання освіти, науки і бізнесу;
- формування громадської думки;
- розвиток і поширення іміджу ВНЗ.

Комунікаційна робота ВНЗ — це сукупність процесів передачі інформації, даних про стан і діяльність вищого навчального закладу, які піддаються цілеспрямованому регулюванню й використовуються для формування іміджу, ідеологічного впливу на оцінки, думки й поведінку споживачів освітніх та наукових послуг.

Комунікація характеризується:

- розосередженістю аудиторії;
- наявністю широкомасштабних комунікаційних процесів;
- одночасним передаванням даних різними каналами масового інформування: через книжки, газети й журнали, радіо- й телетрансляції, комп'ютерні мережі тощо;
- принциповою доступністю всім охочим. [37].

У системі відносин між вищим навчальним закладом та ринком праці, абітурієнтами, батьками, суспільством в цілому процеси масової комунікації є

найважливішим засобом впливу суб'єкта на об'єкт, а також елементом реалізації процесу управління ВНЗ. Роль масової комунікації полягає в наданні суб'єктам процесу можливості здійснювати спрямований вплив на масову свідомість. Усі ті відомості, які поширюються в суспільстві, є інформаційним впливом на масову свідомість, на уявлення, думки й поведінку людей.

Цільовий вплив охоплює в першу чергу випускників шкіл, гімназій, ліцеїв, освітніх закладів середньої професійної освіти. З цією метою друкується реклама в довідкових виданнях і в мережі Інтернет, організуються та проводяться «Дні відкритих дверей», виставкові заходи тощо. У рекламних матеріалах містять інформацію про освітній заклад, факультети, спеціальності та спеціалізації, напрями підготовки (на рівень бакалавра, спеціаліста, магістра, аспіранта, докторанта), вартість навчання, можливості працевлаштування студентів та ін.

Комунікації вищого навчального закладу передбачають обмін інформацією між навчальним закладом та його цільовими аудиторіями. Зміст та форма передачі інформації повинні бути сплановані для задоволення потреб в першу чергу цільової аудиторії та досягнення цілей навчального закладу. Процес організації комунікації вищих навчальних закладів з ринком використовує цілий комплекс засобів різноманітних функцій і різну внутрішню структуру. Всі разом вони складають складний комплекс promotion-mix, що означає систему просування товару, яку прийнято називати як інтегрована система маркетингової комунікації. [37].

Комунікаційні структури повинні відповідати наступним критеріям:

- наукова обґрунтованість – відображає, чи враховується дія економічних законів;
- цілісність – передбачає обов'язковість виконання всього комплексу процедур і специфічних функцій;
- точність – характеризує, чи забезпечується адекватність і достовірність результатів;
- результативність – показує, чи підвищується рівень і ефективність діяльності комунікаційної структури;

— економічність – характеризує, чи досягається ефект від діяльності підрозділу, чи є вигоди для всієї організації.

Таким чином, комунікації в організації управління структурними підрозділами вищого навчального закладу можна розглядати у двох аспектах:

— як процес інформаційної взаємодії між співробітниками при виконанні своїх функціональних обов'язків;

— як функцію управління, яка дозволяє за допомогою правильно організованої передачі інформації забезпечити ефективність співпраці всіх підрозділів та їх взаємодію.

Освітнім закладам потрібні ефективні комунікації з їх ринками та цільовими аудиторіями. Для забезпечення успішної комунікативної діяльності ВНЗ необхідні ефективні комунікаційні програми, що вимагають високого професіоналізму комунікаторів та характеризуються певними ознаками, а саме:

- скоординованість (аналіз різноманітних цільових аудиторій для визначення їх комунікаційних потреб, визначення бажаної реакції від кожної групи, розроблення комунікаційних програм для кожної цільової аудиторії);

- єдиний логотип у ЗМІ та інших комунікаціях для збільшення впізнання та ідентифікації;

- персоналізація комунікацій за допомогою комп'ютеризованих списків адрес під час роботи приймальної комісії;

- об'єктивність (правдивість) комунікативної інформації, адже її розбіжності із дійсністю скоро виявляться абітурієнтами чи студентами. Крім того, ефективні комунікаційні програми постійно потрібно переглядати для перевірки якості та послідовності змісту і стилю та оновлювати відповідно до змін у діяльності навчального закладу чи життєдіяльності студентів у ньому. [38].

Потоки інформації, що слугують основою комунікацій, дозволяють закладу вищої освіти реагувати на зміни у зовнішньому та внутрішньому середовищі.

Характеризуючи організаційні комунікації, слід зазначити, що їх можна поділити на декілька великих типів. У залежності від суб'єктів комунікаційного процесу комунікації умовно поділяються на внутрішні та зовнішні. Внутрішні комунікації дозволяють здійснювати обмін інформацією в середині закладу вищої освіти, а зовнішні комунікації є обміном інформацією між закладом вищої освіти та зовнішнім середовищем. Вони здійснюються від імені закладу вищої освіти як юридичної особи уповноваженим органом чи представником. Зовнішнє середовище, відіграючи значну роль в діяльності закладу вищої освіти, чинить значний вплив на його функціонування та ефективність.

Аналізуючи комунікації закладу вищої освіти, варто назвати їх основні напрями:

1. *Комунікації між організацією та її зовнішнім середовищем:* комунікації між споживачами через відділ маркетингу; комунікації з податковими органами; комунікації з органами влади; комунікації в результаті пошуку постачальників.

2. *Комунікації між рівнями управління та підрозділами:* комунікації між рівнями управління – вертикальні комунікації, які можуть мати напрями: згори-вниз та знизу-вгору; комунікації між підрозділами – горизонтальні комунікації.

3. *Комунікації керівник-підлеглий,* вони належать до комунікацій по вертикалі і складають основну частину діяльності менеджера.

4. *Комунікації між керівником та робочою групою.* Ці комунікації пов'язані з підвищенням ефективності роботи групи.

5. *Неформальні комунікації* – це канал розповсюдження слухів. Інформація передається по каналах слухів набагато швидше, ніж по каналах формальних повідомлень. Дослідники зазначають, що 80-90% слухів щодо організації можуть бути реальними, але не стосуються окремих осіб або занадто емоційної інформації. [38].

Комунікаційна робота відділу кадрів Донецького національного університету імені Василя Стуса є сукупністю процесів передачі інформації, даних про стан та діяльність закладу вищої освіти, що піддаються цілеспрямованому регулюванню і використовуються з метою формування

іміджу, ідеологічного впливу на оцінки, думки та поведінку споживачів освітніх та наукових послуг. Комунікація Донецького національного університету імені Василя Стуса відзначається такими рисами, як: *розосередженість аудиторії; наявність широкомасштабних комунікаційних процесів; одночасна передача даних різними каналами масового інформування: через книжки, газети та журнали, радіо- і телетрансляції, комп'ютерні мережі тощо; принципова доступність всім охочим.*

У системі відносин між закладом вищої освіти та ринком праці, абітурієнтами, батьками, суспільством загалом процеси масової комунікації є найважливішим засобом впливу суб'єкта на об'єкт, а також елементом реалізації процесу управління закладом вищої освіти. Роль масової комунікації головним чином полягає в наданні суб'єктам процесу можливості здійснювати спрямований вплив на масову свідомість. Цільовий вплив насамперед охоплює випускників шкіл, гімназій, ліцеїв, заклади вищої освіти.

Комуникативний процес Донецького національного університету імені Василя Стуса умовно можна змодельовати наступним чином: джерело (відправник – особа або заклад вищої освіти) сприймає певну подію як завдання, певним чином усвідомлюючи (реагуючи). Потім за допомогою певних засобів передає сприйняте одержувачу у вигляді повідомлення, для чого повинні бути підготовлені доступні матеріали, що подаються у певній формі та певному контексті. Переданий зміст одержувач сприймає, однак у зміненому вигляді, тобто сприймає очима відправника. [39].

Реакція, результати, а отже, дієвість комуникативного процесу визначається встановленням та підтриманням контакту між комунікаторами, а його ефективність – ступенем згоди між ними. З цією метою використовується чимало різних комуникативних методів, способів, вираження відчуттів, емоцій, певної організації комуникативного процесу тощо. Комунікації Донецького національного університету імені Василя Стуса передбачають обмін інформацією між навчальним закладом та його цільовими аудиторіями. Зміст і форма передачі інформації зазвичай сплановані в першу чергу для задоволення

потреб цільової аудиторії та досягнення цілей навчального закладу. Процес організації комунікації закладу вищої освіти з ринком передбачає використання цілого комплексу засобів різноманітних функцій та різну внутрішню структуру. В сукупності вони формують складний комплекс *promotion-mix*, який означає систему просування товару, тобто інтегровану систему маркетингової комунікації. [39].

Структура інтегрованої системи маркетингової комунікації закладу вищої освіти містить: рекламу, профорієнтаційну роботу, виставкову діяльність, паблік рілейшенз (PR), особисті комунікації, інтернет. Ця система є комплексом взаємопов'язаних елементів, завдяки яким здійснюється збір, обмін, обробка та передача інформації від одного суб'єкта до іншого. Взаємодія, уніфікація необхідної інформації, джерел та потрібних людей впливає на ефективність діяльності закладу вищої освіти. В кінцевому результаті успішні комунікації підводять споживача до вибору закладу вищої освіти, а також відбувається трансформація комунікативної ефективності в економічну.

Різні види комунікацій, що здійснюються в Донецькому національному університеті імені Василя Стуса, умовно можна поділити на дві основні групи: зовнішні та внутрішні. Зовнішні комунікації спрямовуються на налагодження ефективних зв'язків закладу вищої освіти з соціумом, з яким він взаємодіє, в якому він функціонує та розвивається (відділами та управліннями освіти районних, міських (обласних) держадміністрацій); Міністерством освіти; законодавчими та контролюючими структурами (контрольно-ревізійне управління, санепідстанція тощо); благодійними фондами, (спонсорськими організаціями; вітчизняними та зарубіжними середніми та вищими навчальними закладами; позашкільними навчальними закладами (центрами дитячої творчості, відділеннями МАН); науково-дослідними інститутами та центрами; економічними та комерційними структурами (банками, промисловими підприємствами, фірмами); профспілковими організаціями: батьками студентів тощо). [39]

Конкретними організаційними формами здійснення зовнішніх комунікацій в Донецькому національному університеті імені Василя Стуса є: ділові зустрічі; переговори (усні чи за допомогою факсового та комп'ютерного зв'язків); розмови телефоном; службові записки; звіти; нормативні та законодавчі документи; угоди тощо.

Загалом використання напрямів PR-комунікацій у сфері освітніх послуг має на меті налагодження відкритого діалогу та гармонізацію відносин між закладом вищої освіти та громадськістю. Такий діалог сприяє підвищенню популярності й створенню ефективного іміджу закладу вищої освіти. Під ефективним іміджем розуміється цілеспрямований імідж, що відповідає очікуванням та запитам споживачів освітніх послуг. Для популярних закладів вищої освіти завдання побудови ефективного іміджу нескладне, тому що придбаний раніше імідж допомагає витримувати конкуренцію. Водночас у зовнішньому середовищі можуть виникнути зміни, які здатні змінити розстановку сил та вплинути на імідж закладу вищої освіти. За таких умов з метою формування концепції ефективного іміджу необхідна концептуальна система заходів, основними положеннями якої є: проведення інформаційної політики, формування компетентної громадської думки про заклад вищої освіти, організація зв'язків з громадськістю (PR) з використанням комунікативного менеджменту.[40]

Проведення інформаційної політики в Донецькому національному університеті імені Василя Стуса за допомогою організації потоку новин про всі сфери діяльності закладу вищої освіти в засобах масової інформації забезпечує інформування населення про діяльність навчального закладу, завоювання інформаційного простору на рівні освіти, регіону, країни тощо. При цьому навчальний заклад у своїй діяльності дотримується принципів відкритості та конструктивного діалогу, забезпечує повноту інформації про стратегію і плани розвитку та поточні події, прагне до найбільш повного задоволення суспільної потреби в достовірній інформації. При здійсненні інформаційної політики враховуються і так звані інформаційні потоки та відбувається управління ними.

При реалізації інформаційної політики в Донецькому національному університеті імені Василя Стуса враховується роль носіїв інформації та інформаційних партнерів, а також специфіка сучасного інформаційного простору. [40]. Так, зокрема, носії інформації можуть бути внутрішніми та зовнішніми. Наприклад, для інформування колективу закладу вищої освіти використовується корпоративна газета, інтернет-сайт, наради, конференції. Для інформування на зовнішньому рівні – постійно функціонуючий сайт за умови регулярного оновлення його сторінок. В якості інформаційних партнерів залучаються ЗМІ; інформація повідомляється через прес-релізи, брифінги, прес-конференції, виступи керівництва закладів вищої освіти у пресі, в новинних програмах ТБ. При цьому має місце узгодженість дій, а також враховується специфіка інформаційного простору, що розглядається як поле взаємодії суб'єктів інформаційного процесу, де суб'єктом здійснюється подання спеціально організованої інформації з метою впливу на свідомість об'єкта. [40].

Включення чинника формування громадської думки в концепцію створення ефективного іміджу Донецького національного університету імені Василя Стуса зумовлене: специфікою функцій (освітня, виховна, директивна тощо), залученням громадської думки в управлінський процес; розумінням громадської думки, а також взаємозв'язком та доповнюваністю інформаційної політики й громадської думки. Виходячи з настанови, що формування громадської думки закладу вищої освіти – це складний процес, компонентами якого є правильний вибір цілей та завдань, виявлення та аналіз індивідуальних думок, система раціонального планування і прогнозування, організація взаємодії керівництва закладу вищої освіти й аудиторії (студенти, співробітники, населення) та ефективний контроль, можна констатувати, що є необхідною модель формування громадської думки. Ця модель потрібна для того, щоб: навчитися керувати процесом формування громадської думки у закладі вищої освіти та визначати найкращі способи управління за даних умов; прогнозувати прямі та непрямі наслідки реалізації заданих способів і форм впливу на суб'єкт

громадської думки з виходом на бажаний результат; задавати технологічність процесу формування громадської думки. [40].

Сфера відповідальності за формування іміджу Донецького національного університету імені Василя Стуса перетинається з основною метою та об'єктом діяльності – зв'язками з громадськістю. Напрямами діяльності у сфері зв'язків закладу вищої освіти з громадськістю є: загальнофірмова комунікація (внутрішня та зовнішня). Внутрішня комунікація – це навіювання працівникам віри у свій заклад, почуття гордості та відповідальності за нього. Для її забезпечення передбачається:

- проведення заходів, які сприяють встановленню добрих стосунків між керівниками та підлеглими (вечірки, екскурсії, спортивні змагання тощо);
- надання соціальних пільг працівникам (додаткові відпустки, пенсійні та страхові внески тощо);
- видання газети закладу вищої освіти;
- підтримка зразкової репутації керівників у суспільстві;
- матеріальні стимули (премії, висока заробітна плата тощо), що впливають на зміцнення позицій закладу вищої освіти.

Зовнішня комунікація – це створення образу закладу вищої освіти в очах громадськості міста, країни; при цьому важливо своєчасно повідомляти громадськість про досягнення закладу, його успіхи. Для її забезпечення передбачається:

- встановлення та підтримка стійких зв'язків із засобами масової інформації. Цю роботу спрямовано на розміщення відомостей і матеріалів пізнавально-інформаційного характеру в засобах масової інформації для привертання уваги до певних осіб, освітніх послуг;
- освітня пропаганда. Здійснення різноманітних заходів, що забезпечують впізнаваність певної послуги. Для цього використовуються: редакційне місце або час у різних ЗМІ, доступні потенційним абітурієнтам. Інколи вона спрямована на конкретну освітню спеціальність. Пропагується саме закладу вищої освіти;

- лобізм. Діяльність, спрямована на ухвалення, неухвалення чи скасування законодавчих актів та урядових рішень через вплив на думку законодавців і чиновників;
- консультування. Розроблення та подання настанов, пропозицій, рекомендацій щодо питань суспільної значущості, становища, репутації та образу закладу вищої освіти. [41].

Для ефективної роботи PR-технологій в Донецькому національному університеті імені Василя Стуса використовуються різні канали комунікації – ініціюється та розміщується інформація не лише у пресі, а й на телебаченні та в інтернеті; відбувається робота з різними цільовими аудиторіями – діловими, громадськими тощо, орієнтованими на питання освіти; організовуються заходи, орієнтовані на цільові аудиторії, наприклад, «один день у закладі вищої освіти», або участь у майстер-класі, організація прес-туру та екскурсії по закладу вищої освіти.

Інформаційне висвітлення та організація спеціальних заходів у Донецькому національному університеті імені Василя Стуса є частиною комплексного підходу до вирішення комунікативних завдань, в основі якого лежить визначення і дотримання місії закладу вищої освіти. Зміст місії є інструментом формування загального образу вищої школи нарівні зі створенням візуального стилю, а також організацією представництва в інтернеті. [41].

2.2 Система комунікацій у внутрішньому середовищі закладу вищої освіти

Б. Адамс виокремлює такі види комунікацій:

- формальні – обговорення між співробітниками робочих процесів, які визначаються організаційною структурою установи;
- неформальні – спілкування між колегами, як між друзями, на різну тематику, чи канал поширення пліток;
- вертикальні – зверху вниз і знизу вгору;

- горизонтальні – обмін інформацією між управліннями та відділами. [42].

Комунікаційні процеси, в яких бере участь Донецький національний університет імені Василя Стуса, можуть бути розділені на два великих види: комунікація всередині організацій і зовнішня комунікація. Розглянемо більш детально перший вид комунікаційного процесу.

Загалом внутрішні комунікації забезпечують взаємодію між усіма ланками управлінського процесу, управліннями, відділами, окремими особами в процесі особистих контактів або обміну документацією. Внутрішні комунікації поділяють на три групи: *друковані; технологічні; персональні*. [43].

Внутрішні комунікації забезпечують взаємодію працівників закладу вищої освіти між собою (налагодження ефективних зв'язків між членами управлінської команди (визначення стратегічних і тактичних завдань, розподіл функціональних обов'язків, координація дій):

- організація ефективного навчально-виховного процесу;
- обмін педагогічним досвідом, підвищення професійного рівня членів трудового колективу ЗВО;
- оцінка ефективності діяльності працівників закладу вищої освіти (при здійсненні різних видів контролю, проведенні атестації);
- організація спільної діяльності викладачів та студентів;
- налагодження стосунків між членами ЗВО;
- створення умов для фахового і творчого зростання працівників;
- формування традицій закладу вищої освіти тощо. [44].

Внутрішні комунікації, як і зовнішні, можна розділити на види: вертикальні комунікації, які здійснюються між членами закладу вищої освіти, що мають різний «управлінський статус»; горизонтальні – комунікації, що здійснюються між членами ЗВО, що займають однакове місце в системі управлінських стосунків. [45].

Внутрішні комунікації - це будь-які комунікації всередині організації. Вони можуть бути усними або письмовими, безпосередніми або віртуальними,

особистими або груповими. Створення ефективних внутрішніх комунікацій всіх напрямків - зверху вниз, знизу вгору і по горизонталі - це одна з основних завдань відділу кадрів в ЗВО.

Мета внутрішньої комунікації відділу кадрів - створення серед співробітників підтримки цілей і політики, що проводиться керівництвом організації. За допомогою комунікації співробітники отримують необхідні знання та мотивацію, стають кращими захисниками і популяризаторами планів університету. Крім того, внутрішня комунікація дозволяє встановити рольові взаємодії і розподілити відповідальність працівників, що важливо в забезпеченні успіху в робочому процесі. За внутрішніми каналами комунікацій передаються два види інформації: що відбувається всередині навчального закладу і яке місце компанія займає у зовнішньому середовищі. [46].

Структура організації визначає напрямки комунікаційних потоків в ній, тобто руху інформації по різних каналах між різними підрозділами та працівниками цієї організації:

- *Вертикальні комунікаційні потоки* йдуть від адміністрації до рядових членів (наказ, постановка задач, переконання, підтримка, контроль, кризова комунікація) і від рядових членів до адміністрації (звіт, запит, службова записка, профспілкові збори). Звернення вгору по інстанції означає підключення до ланцюга команд, представленої в будь-якій схемі організаційної структури вертикальними лініями. Якщо співробітник не може вирішити проблему, то її рішення передається далі - вгору по інстанції. Таким чином, вертикальні лінії організаційних схем служать внутрішніми комунікаційними каналами; [47].

- *Горизонтальні комунікаційні потоки* відбуваються між рівними за рангом членами організації, відносяться до комунікацій з координації, які існують між організаційними підрозділами (виробнича комунікація, обмін думками, збори, кава-брейк, бесіда після роботи і ін.). Горизонтальні канали комунікації покликані вирішувати такі завдання: передачу інформації не директивного, але дорадчого характеру; уточнення цілей і завдань підрозділів виходячи з конкретних ситуацій в кожному з підрозділів; взаємодія фахівців різного

профілю з різних підрозділів організації, що сприяє комплексному вирішенню поставлених вищим керівництвом завдань; формування неформальної структури організації, рішення задач, які неможливо вирішувати в офіційних рамках; [48].

- *Зовнішні комунікаційні потоки* - це будь-які повідомлення, пов'язані з перетином умовної межі організації (реклама і корпоративна реклама, статті, інтерв'ю, виставки, презентації, прес-конференції, прес-релізи, кризова комунікація).

Внутрішні комунікації повинні допомагати співробітникам розуміти корпоративну культуру, цілі і цінності компанії, в даному випадку університету. У внутрішніх комунікаціях важливі суть, канали і навіть форма інформації. [49].

Форма дуже сильно впливає на повідомлення: один і той же факт можна повідомити так, що він буде сприйнятий в негативному або позитивному ключі, залишить людей байдужими або, навпаки, викличе інтерес. Якщо компанія вкладає величезні кошти і застосовує самі передові технології для того, щоб забезпечити собі PR - тобто відносини з громадськістю, - то вона повинна так само уважно ставитися і до внутрішніх відносин. Найскладніше - знайти потрібний баланс між комунікацією "зверху вниз" і "знизу вгору", щоб демократичність і сприйнятливість менеджменту до нових ідей не впливала на старанність і дисциплінованість співробітників. [50].

До складу каналів внутрішніх комунікацій входять:

- Внутрішній веб-сайт (Інтранет);
- Корпоративна преса, корпоративні бюлетені, прес-кліппінг і інші неперіодичні друковані матеріали;
- Внутрішня e-mail -розсилка;
- Дошки оголошень, розпорядження керівництва, різні види звітності; [51].
- Регулярні зібрання, в тому числі:
 - а) неформальні зустрічі, на яких працівники можуть поговорити з начальством безпосередньо;
 - б) віртуальні зборів з використанням онлайн-платформ типу Second Life,
 - в) конференції і ін. [52].

До комунікативних засобів відносять : взаємодія співробітників (усні, групові бесіди); зустрічі з першими особами компанії, зборів; заходи, що проводяться підрозділами; семінари, програми навчання; корпоративний кодекс і базу знань; корпоративні заходи, конкурси, акції; внутрішні екскурсії; управління чутками і ін. [53].

Для встановлення комунікацій з працівниками керівництво використовує різні підходи.

1. **Націлений підхід:** комунікація встановлюється з чітко визначеною цільовою аудиторією. Чим точніше складено повідомлення, тим більша ймовірність його вірного сприйняття. Головна помилка - це вважати, що інформація передається тільки за допомогою слів і що одержувач пасивно приймає повідомлення. [53], [54].

2. **Круговий підхід:** комунікація встановлюється за допомогою хороших людських відносин, а також роботи над тим, щоб кожен співробітник отримував задоволення від роботи. Таку систему можна встановити тільки через тривалі і відкриті дискусії. Передбачається, що сутність комунікації - в досягненні взаєморозуміння. Проблеми виникають через неправильне уявлення про те, що розуміння призводить до угоди і в цьому єдина мета спілкування. Необов'язково розділяти точку зору людини, щоб поважати її і враховувати в роботі.

3. **Підхід активної дії:** комунікація встановлюється за допомогою практичних дій, в яких необхідні розуміння і інтуїція. Підхід ґрунтується на тому принципі, що комунікація - це координація значень, розуміння загальних правил і розпізнавання образів. [55], [56].

Забезпечення ефективних внутрішніх комунікацій не вимагає істотних витрат, однак це досить трудомісткий процес, необхідний для розвитку компанії і підтримки її працездатності.

Умовно всі інструменти внутрішнього PR закладу вищої освіти можна розділити на 4 групи.

1. **Інформаційні.** Сюди належать ЗМІ (зовнішні та внутрішні). Зовнішні – якщо інформацію про заклад вищої освіти публікує якесь поважне ЗМІ, нехай

працівники про це дізнаються, отримають примірник цього журналу чи побачать його на нараді тощо.

2. *Аналітичні.* До них належать: Зворотний зв'язок. Варто придумати систему, яка б його забезпечувала у закладі вищої освіти. Це може бути окрема електронна скринька з листами керівництву чи щоквартальні зустрічі віч-на-віч. Не важливо, який формат. Головне, аби він працював. Анкетування. Слід завжди бути в курсі потреб працівників. І найкращий спосіб про них дізнатись – запитати прямо. Варто розробити анкетування про те, що в умовах праці ваші працівники хотіли б покращити.

3. *Комунікативні.* Сюди належать всі активності, що зазвичай хибно вважають внутрішніми комунікаціями, зокрема: корпоративні свята, змагання, тренінги, вітання тощо. Відчуваючи позитивні емоції, в людей виростає більша довіра та прихильність до тих, з ким вони їх пережили.

4. *Організаційні.* Варто пам'ятати, що найголовнішу інформацію працівники все-таки повинні отримувати від свого керівника. Тому головні меседжі, формат спілкування, ключові слова для нього повинна підбирати людина, яка розуміється на цьому – менеджер з внутрішніх комунікацій. [57], [58].

Разом із використовуваними видами та типами комунікацій в Донецькому національному університеті імені Василя Стуса існують ще й комунікаційні технології управління персоналом. Комунікації в системі підбору, відбору та найму персоналу закладу вищої освіти включають в себе:

- комунікації з кандидатами в процесі підбору на вакантну посаду;
- співбесіди в процесі відбору;
- аналіз як усної, так і письмової інформації про кандидатів;
- контакти з керівниками підрозділів, в які здійснюється підбір персоналу. [59].

Комунікації в процесі формування корпоративної культури Донецького національного університету імені Василя Стуса включають:

- інформування про цілі та цінності закладу вищої освіти;

- забезпечення зворотного зв'язку з працівниками навчального закладу. [60].

Комунікації при управлінні конфліктами в Донецькому національному університеті імені Василя Стуса включають в себе: *контакти з учасниками конфлікту; аналіз інформації, отриманої з різних джерел про суть конфлікту; діалоги в процесі вирішення конфлікту.*

Комунікації в процесі взаємодії керівників і підлеглих здійснюються у такі способи: *інформування підлеглих керівництвом закладу вищої освіти; звітування підлеглих; обговорення завдань та шляхів їх вирішення; розробка правил дистанційного та особистого спілкування.* [61], [62].

Комунікації управління поведінкою персоналу закладу вищої освіти включають: *розробку письмових стандартів поведінки персоналу закладу вищої освіти; бесіди з порушниками дисципліни; підготовку документів за результатами дисциплінарного аналізу.* [62].

2.3 Механізми функціонування програм-парсерів у відділі кадрів Донецького національного університету імені Василя Стуса

Загалом існує два основні механізми функціонування програм-парсерів в реалізації пошуку. Перший полягає в прямому пошуку серед існуючих систем зберігання даних (HTML файли, бази даних... залежно від типу сайту). Цей механізм дуже повільний, оскільки типи даних, що зберігаються і до яких ви маєте доступ, як правило, не передбачені для цієї мети.

Другий механізм, полягає в створенні способу зберігання даних, який відповідатиме конкретним вимогам пошуку. Наприклад, оцінка для кожного слова в статті може зберігатися таким чином, щоб бути легко доступною знову і щоб дозволяти отримати загальний бал пошуку додаванням.

Перевага тут у тому, що пошук відбувається дуже швидко - майже так само швидко, як і будь-які інші розрахунки для сторінки, які ми могли б зробити. [63].

Незручність в тому, що цей механізм вимагає фази для побудови інформації, призначеної для зберігання; ця фаза відома як крок індексації. Індексація має певну вартість в плані ресурсів (часу для її виконання і дискового простору), а також вводить невеликі затримки в часі між додаванням або модифікацією змісту і тим, коли цей доданий чи модифікований зміст відбиватиметься в результатах пошукової системи. З іншого боку, у разі парсерування, ми змушені використовувати PHP і MySQL, як ми це робимо для іншої частини програмного забезпечення, що не допоможе нам досягти високої продуктивності пошуку, не тільки з точки зору швидкості, але й з боку релевантності та різних інших удосконалень (індексування документів, зовнішніх щодо нашого сайту; створення семантичних полів, що дозволяють більш точно налаштовувати пошук; і т.д.).

Пошукова машина, як парсер і інші являють собою потужні інформаційно-пошукові системи, розміщені на серверах вільного доступу, спеціальні програми яких безперервно в автоматичному режимі сканують інформацію мережі на основі заданих алгоритмів, проводячи індексацію документів. У подальшому пошукові машини надають користувачеві на основі створених баз даних доступ до розподіленої на вузлах мережі інформацію через виконання пошукового запиту в рамках власного інтерфейсу.

Освоєння інтерфейсу пошукової машини є прекрасним фундаментом для роботи з іншими пошуковими системами. [64].

Як правило, пошукові машини забезпечують інтерфейс типу меню, за допомогою якого користувач може скомпонувати запит на пошук інформації, використовуючи ключові слова та /або фрази і логічні зв'язки І-АБО-НЕ. Більшість машин пошуку знаходять величезну кількість релевантних сторінок за запитом користувача. Кожен знайдений документ зазвичай ранжирується по мірі його кореляції із запитом. Релевантність кожного документа оцінюється за допомогою різних технологій, наприклад обліку частоти появи на сторінці шуканих слів. Деякі пошукові механізми використовують додатково інші

фактори, такі як частота відвідування сторінки та /або близькість розташування один до одного шуканих термінів. [65].

Пошуковий сервер є найважливішим елементом всієї системи, оскільки від його алгоритмів функціонування, безпосередньо залежить якість та швидкість пошуку.

Пошуковий сервер працює наступним чином:

- Отриманий від користувача запит піддається морфологічному аналізу. Генерується інформаційне оточення кожного документа, що міститься в базі (зворотній індекс).
- Отримані дані передаються в якості вхідних параметрів до спеціального модулю ранжирування. Відбувається обробка даних по всіх документах, в результаті чого, для кожного документа обчислюється власний рейтинг, що характеризує релевантність запиту, введеного користувачем, і різних складових цього документа, що зберігаються в індексі пошукової системи.
- Залежно від вибору користувача цей рейтинг може бути скориговано додатковими умовами (наприклад, так званий «розширений пошук»).
- Далі генерується сніппет, тобто, для кожного знайденого документа з таблиці документів витягуються заголовок, коротка анотація, найбільш відповідна до запиту і посилання на сам документ, причому знайдені слова виділено грубішим шрифтом.
- Отримані результати пошуку передаються користувачеві у вигляді SERP (*Search Engine Result Page*) - сторінки видачі пошукових результатів. [66].

Як видно, всі ці компоненти тісно пов'язані один з одним і працюють у взаємодії, утворюючи чіткий, достатньо складний механізм роботи пошукової системи, що вимагає величезних витрат ресурсів.

Найважливішим завданням та механізмом функціонування пошукових систем є виявлення та індексація релевантних сторінок. Тому дуже важливо своєчасно оновлювати інформацію в базах даних, як нову, так і вже існуючу, щоб вона відповідала поточним запитам.

Пошукові системи заносять в базу для індексації мільйони сторінок в день. І тут відразу виникає проблема їх частоти оновлення. Адже час, коли спайдер знову прочитає вже існуючу в базі сторінку, може бути досить великим, а інформація може кілька разів на сторінці змінитися. Пошукові системи не можуть постійно стежити за оновленнями інформації веб-сторінки, ось саме цей фактор є основною проблемою релевантності. [67].

Якщо зовні всі пошукові системи здаються схожими, то завжди у них будуть відрізнятися результати видачі пошукових запитів, що залежать від індексації, частоти оновлення баз даних і алгоритмів вибірки.

В нашому випадку, використання парсингу у відділі кадрів Донецького національного університету імені Василя Стуса є багатоплановим. Розподіл навчального навантаження на кафедрі починається з розбору файлу PDF з навчальним навантаженням, який надсилають методисту кафедри. Методист, керуючись положенням про нормування робочого часу, виконує розподіл навчального навантаження між викладачами. Після розподілу навантаження отримуємо звіт про навантаження для кожного викладача та архів розподілу навантаження, зокрема, базу даних, в якій буде зберігатися інформація з навантаженням кожного викладача. Контекстна діаграма структурно-функціональної моделі описує модель процесу розподілу навантаження в узагальненому вигляді. [68]

«Формування навантаження викладачів кафедри» декомпозиється на два процеси: заповнення планового навантаження та розподіл навантаження між викладачами. Блок «Заповнення планового навантаження» на вході має PDF-файл навантаження. На виході ми отримуємо дані навантаження в базу даних. Управління процесом здійснюється згідно положення про нормування робочого часу, а також штату кафедри. Механізмами процесу виступають програмне забезпечення і методист кафедри. [69], [70].

Процес «Заповнення планового навантаження» виконується в чотири етапи, які в діаграмах будуть представлені відповідними на чотири процеси – саме: парсинг, перевірка коректності парсингу, за потреби редагування

отриманої інформації та її завантаження в базу даних. Вхідною інформацією для парсингу є PDF-файл з навантаженням. На виході ми отримуємо розпарсені дані, які містять конкретні значення відповідних комірок таблиці PDF-файл. Механізмом процесу виступає програмне забезпечення «Перевірка коректності парсингу», яке на вході має розпарсені дані, PDF-файл з навчальним навантаженням і відредаговані дані.

Управління здійснюється згідно положення про нормування робочого часу і штатом кафедри. Механізмом є програмне забезпечення і методист кафедри. На виході отримуються коректні дані, або некоректні дані. На вхід процесу «Редагування» подаються некоректні дані. Механізмом є програмне забезпечення і методист. На виході отримуються відредаговані дані. «Завантаження в БД» на вході має коректні дані. Механізмом процесу виступає програмне забезпечення, а на виході отримуються дані навантаження у базі даних. На вхід процесу «Розподіл навантаження» подаються дані навантаження в БД. На виході отримується звіт про навантаження викладача і архів розподілу навантаження. Управління процесом здійснює штат кафедри, а механізмом є програмне забезпечення і методист кафедри.

Розподіл навантаження декомпозиється на чотири процеси, а саме вибірку інформації з бази даних, закріплення викладача, збереження в базі даних, формування звітності для викладача. «Вибірка інформації з бази даних» на вході має дані навантаження в базу даних. На виході отримуються обрані дані. Механізмом є програмне забезпечення. На вхід процесу «Закріплення викладача» подаються обрані дані. На виході отримують закріплені викладачі. Управління здійснюється штатом кафедри. Механізмом є програмне забезпечення і методист кафедри. «Збереження в базі даних» на вході має закріпленого викладача. На виході архів розподілу навантаження. Механізмом є програмне забезпечення. «Формування звітності для викладача» на вході має архів розподілу навантаження. На виході отримується звіт про навантаження викладача. Механізмом є методист кафедри. [71], [72].

Висновки до розділу 2

Отже, сучасні мультимедійні комп'ютерні програми та інформаційно-комунікаційні технології відкривають доступ до нетрадиційних джерел інформації - електронних підручників, освітніх Web-сайтів.

Комунікаційна робота ВНЗ — це сукупність процесів передачі інформації, даних про стан і діяльність вищого навчального закладу, які піддаються цілеспрямованому регулюванню.

Комунікація Донецького національного університету імені Василя Стуса відзначається такими рисами, як: *розосередженість аудиторії; наявність широкомасштабних комунікаційних процесів; одночасна передача даних різними каналами масового інформування тощо.*

Різні види комунікацій, що здійснюються в Донецькому національному університеті імені Василя Стуса, умовно можна поділити на дві основні групи: зовнішні та внутрішні.

Використання парсингу у відділі кадрів Донецького національного університету імені Василя Стуса є багатоплановим. Розподіл навчального навантаження на кафедрі починається з розбору файлу PDF з навчальним навантаженням, який надсилають методисту кафедри. Методист, керуючись положенням про нормування робочого часу, виконує розподіл навчального навантаження між викладачами.

«Формування навантаження викладачів кафедри» декомпозиується на два процеси: заповнення планового навантаження та розподіл навантаження між викладачами

Розподіл навантаження декомпозиується на чотири процеси, а саме вибірку інформації з бази даних, закріплення викладача, збереження в базі даних, формування звітності для викладача. Механізмом є програмне забезпечення. «Формування звітності для викладача» на вході має архів розподілу навантаження. На виході отримується звіт про навантаження викладача. Механізмом є методист кафедри. [72].

РОЗДІЛ 3

КОНЦЕПЦІЯ ФОРМУВАННЯ КОМУНІКАЦІЇ З ВИКОРИСТАННЯМ ПРОГРАМ-ПАРСЕРІВ У ВІДДІЛІ КАДРІВ

3.1 Проблеми впровадження методу парсингу в закладах вищої освіти в напрямі забезпечення інформацією відділу кадрів

В умовах прискорених темпів інформатизації, все більшу вагу набуває проблема пошуку потрібної інформації. Разом зі збільшенням об'єму інформаційних ресурсів постійно зростають вимоги до сучасних пошукових систем. Серед головних з них виділяють швидкість пошуку, релевантність, ресурсні затрати, якість отриманих даних та гнучкість пошукових процесів.

На сьогоднішній день наявність власного інтернет-сервісу є зручним і перспективним напрямом в освіті. Є можливість створення інтернет-сервісу самостійно, але якщо не вистачає компетентності в програмуванні, то можна створити інтернет-сервіс з надання інформаційно-освітніх послуг за допомогою парсингу. Програма-парсер здатна «парсити» інформацію з необхідної категорії. Використовуючи програму-парсер можна самостійно створювати будь-які сервіси. Парсинг сайтів є ефективним рішенням для автоматизації збору і зміни інформації. [73]. У порівнянні з людиною, програма-парсер здійснює 4 кроки при опрацюванні масиву даних:

1. швидко обійде тисячі веб-сторінок;
2. обережно відокремить технічну інформацію від «людської»;
3. безпомилково відбере потрібне і відкине зайве;
4. ефективно упакує кінцеві дані в необхідному вигляді.

Результат (будь-то база даних чи електронна таблиця, чи інформація про працівників та студентів закладу вищої освіти) потребує подальшої обробки. Втім, подальші маніпуляції з зібраною інформацією вже до теми парсингу не належать. [73]. Як правило, потрібно «парсити» не одну сторінку сайту-донора,

а безліч, може навіть і все. У цьому випадку після проходження перших трьох кроків у алгоритм програми-парсера повинен бути закладений перехід на наступну сторінку бази, щоб і з неї вилучити весь необхідний матеріал.

В нашому випадку, одним із ефективних способів вирішення проблеми є розробка та впровадження спеціалізованих пошукових систем, але цьому передуює проблема самого впровадження. Адже коли появляється, щось нове, то відбувається протидія програмі з боку робітників, працюючих у відділі кадрів.

Обхід всіх потрібних сторінок сайту забезпечується різними способами:

- По-перше, обробляючи чергову сторінку, програму-парсер можна навчити не тільки отримувати необхідні дані, але і заносити в свою базу даних всі внутрішні посилання, що зустрічаються на шляху. Звертаючись до свого сховища лінків, програма послідовно відвідує сторінки сайту, до тих пір поки не обійде їх всіх.
- По-друге, при первинному аналізі сайту найчастіше можна простежити логіку формування url для сторінок. І потім, генерувати адреси відповідно до виявлених закономірностей.
- По-третє, деякі програми-парсери розраховані, як не дивно, на «ручний» обхід веб-ресурсу. Користувач, клікаючи по посиланнях, сам вирішує які сторінки відвідувати, які ні. А програма у фоновому режимі запам'ятовує необхідні дані. Зрозуміло, поєднувати різні методи ніщо не заважає. [74].

Є кілька варіантів обмежень, які можуть ускладнити роботу парсера при використанні його у відділі кадрів:

- *За user-agent.* Це запит, у якому програма повідомляє сайт про себе. Парсери банять багато веб-ресурсів. Однак у налаштуваннях дані можна змінити на YandexBot або Googlebot та надсилати правильні запити.
- *За robots.txt,* в якому прописано заборону для індексації пошуковими роботами Яндекс або Google (ними ми представилися сайту вище) певних сторінок. Необхідно встановити в налаштуваннях програми ігнорування robots.txt.

- *За IP- адресою*, якщо з неї протягом тривалого часу надходять на сайт однотипні запити. Рішення – використовувати VPN.
- *По капчі*. Якщо дії схожі на автоматичні, виводиться капча. Навчити парсери розпізнавати конкретні види досить складно та дорого.

Використання спеціалізованих пошукових систем, орієнтованих на пошук інформаційних даних у заданій вузькій області, дозволяє значно зменшити об'єм пошукових ресурсів. Іншою перевагою використання спеціалізованих пошукових систем є можливість використання пошукових фільтрів. Усе це дозволяє значно збільшити швидкість пошуку потрібних даних, підвищити релевантність, а також зменшити ресурсні затрати на виконання самого пошукового процесу.

Швидкість пошуку - це параметр, який характеризує проміжок часу від отримання пошуковою системою запиту до виведення нею результатів пошуку. Швидкість пошуку в спеціалізованих пошукових системах вища, оскільки досить велика частина інформації, яка не належить до даної категорії, відсіюється ще до початку пошуку, інша частина надлишкової інформації відсіюється внаслідок використання пошукових фільтрів.

Релевантність пошуку - параметр, який показує, наскільки відповідає початковий запит отриманим результатам пошуку. Чим вища релевантність, тим краще функціонує пошукова система. У стандартних пошукових системах релевантність нижча за рахунок матеріалів з різноманітних категорій даних, які задовольняють пошуковий запит, наприклад, за ключовими словами, але не відносяться за змістом до шуканих даних. [75].

Обсяг інформації у спеціалізованих пошукових системах значно менший, і, зазвичай, усі їх інформаційні ресурси використовуються в стандартних пошукових системах. Дану проблему можна вирішити, спільно використовуючи кілька спеціалізованих пошукових систем. Ресурсні затрати в спеціалізованих пошукових системах нижчі, що дозволяє виконання більшої кількості запитів за фіксовану одиницю часу. Якість отриманих даних у спеціалізованих пошукових

системах забезпечується фільтруванням інформаційних ресурсів, які не відносяться до категорії, за якою проводиться пошук.

Гнучкість пошуку - це параметр, що вказує, наскільки дії користувача можуть вплинути на пошуковий процес з урахуванням додаткових параметрів пошукових операцій. Гнучкість пошуку забезпечується за допомогою використання пошукових фільтрів. Під чітко визначеною цільовою аудиторією розуміємо кваліфікованих користувачів, які здатні здійснювати ефективний пошук шляхом формування точних запитів за чітко визначеними типами даних. Розробка спеціалізованих пошукових систем є актуальною задачею в галузі інформаційних технологій, спрямованою на оптимізацію пошукових процесів та забезпечення їх якісних показників.

У таблиці 3.1. наведені проблеми, які можуть сповільнити або й навіть призупинити впровадження програм-парсерів у відділі кадрів. [76].

Таблиця 3.1 – Проблеми, які можуть сповільнити або призупинити впровадження програм-парсерів у відділі кадрів

Вид проблеми	Опис проблеми	Шляхи вирішення
1	2	3
Небажання вивчати нову програму	У багатьох випадках відбувається переломний перехід від однієї програми до іншої, без урахування особливостей застосованих комп'ютерних технологій. Це призводить до постійних збоїв у роботі комп'ютерної системи та персоналу.	Проведення аналізу існуючої програми й розробка проекту створення та впровадження нової системної програми, яка врахувала би особливості обліку закладу та структуру кадрової комп'ютерної системи, що планується впровадити на підприємстві. Обробляючи чергову сторінку, програму-парсер можна навчити не тільки

1	2	3
		отримувати необхідні дані, але і заносити в свою базу даних всі внутрішні посилення, що зустрічаються на шляху.
<i>Юридична обґрунтованість облікових даних</i>	Хоча основні поняття інформаційного пошуку в законодавстві України, проте розроблених допоміжних нормативно-правових актів, що регулювали би використання програм-парсерів, немає	Необхідно розробити пакет законодавчих актів, які б регулювали порядок пошуку та формування інформації.
<i>Контроль за правами доступу</i>	Користувачі комп'ютерної системи можуть мати доступ до інформації, на користування якої вони не мають права	Необхідно розмежовувати доступ до інформації для відповідних користувачів за рахунок використання різних засобів захисту

Рекомендовані шляхи щодо вирішення та впровадження парсерів у відділі кадрів можна застосовувати також у навчальних цілях, де пов'язаний освітній процес із програмним забезпечення та пошуковими машинами. [76].

3.2 Перспективи використання програм-парсерів у відділі кадрів закладу вищої освіти

Суть інформаційного пошуку у загальному випадку зводиться до того, що пошукова система за певними критеріями обирає з безлічі документів, що знаходяться в базі, такі, які задовольняють інформаційну потребу і відповідають інформаційному запиту (тобто є релевантними). Відповідно до загальних принципів організації інформаційного пошуку, в основі кожного пошукового алгоритму знаходиться модель реалізації, яка використовується для деталізації

пошукової стратегії. Таким чином, можна сказати що формально для пошукового алгоритму вона є математичним представленням, здатним відобразити будь-який релевантний об'єкт в інформаційно-пошуковій системі співвідносно з будь-якими критеріями його використання системою, з метою виконання пошукового завдання. [76].

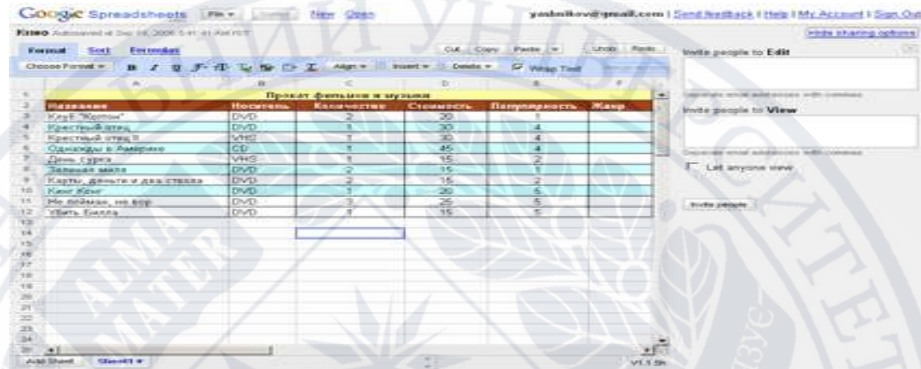
Враховуючи значне накопичення об'ємів інформації в сукупності з прогресуючими темпами зростання її кількості та важливість новацій в області інформаційного пошуку актуальним залишається питання розробки й вдосконалення та впровадження пошукових алгоритмів. Оптимізація інформаційного пошуку з'явилася в період розвитку пошукових систем. У той час пошукові системи надавали велике значення аспектам, якими власники сайтів могли легко маніпулювати: текст на сторінці, ключові слова в мета-тегах та інші внутрішні чинники. Це призвело до того, що у видачі багатьох пошукових систем перші кілька сторінок займали сайти, які були повністю присвячені рекламі. Ідея автоматизованої обробки текстової інформації за допомогою електронно-обчислювальних машин виникла ще на початку XX століття. Розвиток комп'ютерної лінгвістики сприяв інтеграції методів математики (перш за все, статистики та дискретної математики) та лінгвістики для вирішення прикладних завдань аналізу текстової інформації.

Можна навести конкретний приклад програми-парсеру. Це Google Spreadsheet, який може використовувати відділ кадрів Донецького національного університету імені Василя Стуса. [77].

Google Spreadsheet – програма електронних таблиць, що входить до складу безкоштовного веб-програмного пакету програмного забезпечення, пропонованого компанією Google у межах служби Google Drive. Послуга також включає в себе Документи Google і Google Слайди, текстовий процесор і програму презентації відповідно.

Задля роботи з великими масивами документів у Донецькому національному університеті імені Василя Стуса є фахівці, які можуть управляти ними правильно та забезпечувати процвітання закладу вищої освіти в плані

чіткої та достовірної інформації, що стосується студента або працівника університету. Тому відділ кадрів використовуватиме Google Spreadsheet для швидкого пошуку інформації стосовно відпустки, звільнення, прийняття на роботу, відомості про відрахованих студентів або випускників. Найдені дані будуть формуються у електронні таблиці за конкретними ключовими словами, які подані нижче (див. рис. 3.1).



Назва	Тип	Кількість	Ціна	Платіжність	Жанр
Книга "Морган"	DVD	2	30	1	
Книга "Морган"	DVD	1	30	4	
Книга "Морган"	VHS	1	30	4	
Книга "Морган"	CD	1	45	4	
Книга "Морган"	VHS	1	15	2	
Книга "Морган"	DVD	2	15	1	
Книга "Морган"	DVD	2	15	2	
Книга "Морган"	DVD	1	20	5	
Книга "Морган"	DVD	2	25	5	
Книга "Морган"	DVD	1	15	5	

Рисунок 3.1 – Процес роботи Google Spreadsheet з масивом інформації

Розглянемо характеристики даної програми, яка успішно допомагатиме у пошуку тої чи іншої потрібної інформації: [78].

1. Інтерфейс програми-парсеру Google Spreadsheets. У Google Spreadsheets простий і дуже зручний інтерфейс. У ньому активно використовується AJAX (підхід до побудови користувацьких інтерфейсів веб-застосунків), так що більшість дій не вимагає перезавантаження сторінки. Швидкість роботи Google Spreadsheets оцінять навіть користувачі модемного доступу. Якщо ж з'єднання з Мережею буде втрачено, GS запропонує перейти в режим перегляду, при якому будь-які правки вносити зі зрозумілих причин не допускається. [39]. Сам інтерфейс не можна назвати революційним, він виглядає дещо простіше Microsoft Excel. «Перенавчатися» не доведеться. Більш того, новачкові буде простіше починати працювати саме з Google Spreadsheets. Інтерфейс служби зараз представлений лише англійською мовою, але з російською та українською великих проблем немає. Ви можете створювати осередки і стовпці та заповнювати їх кириличними символами. А ось в назві

створених файлів поки припустимі лише латинські літери та цифри. Дуже добре підтримується робота без допомоги миші. Також як в Excel, в GS можна виділити текст і натиснути CTRL + B, CTRL + I тощо, щоб застосувати бажане оформлення. [40]. Коли працюємо з Google Spreadsheets, швидко забувається, що маємо справу з веб-сервісом, а не десктопним додатком. Інтерфейс опрацьований до дрібниць. Наприклад, навіть спливаючі вікна «відкидають» тінь.

2. **Імпорт та експорт даних та файлів.** Google Spreadsheets вміє імпортувати файли в трьох форматах - odf, xls і csv. Але поки що не завжди імпорт проходить гладко. Вбудовані ілюстрації і макроси не імпортуються. Часом оформлення імпортованого файлу може істотно відрізнятись від оригінального, створеного в MS Excel. Максимальний розмір файлу, що імпортується становить приблизно 1 Мб. Що стосується експорту, то створену таблицю користувач може експортувати в кілька форматах - CSV, XLS, HTML, ODS і навіть в PDF. При бажанні дані можна, зрозуміло, і роздрукувати.

3. **Робота Google Spreadsheets.** Google Spreadsheets надає всі основні можливості роботи з електронними таблицями. Користувач може об'єднувати осередки, змінювати накреслення, розмір і гарнітуру шрифту. Допускається запис декількох рядків в одну клітинку. Є кілька варіантів вирівнювання тексту в комірці. При бажанні можна змінити колір шрифту і фон тексту, використовувати кордону. Також як і в Excel, в GS можна створювати в одному файлі кілька таблиць. [41]. Коли відбувається створення таблиці, GS пропонує завжди включення автозбереження. Так що, якщо раптом зв'язок з інтернетом обірветься, дані не будуть втрачені. Остерігатися того, що хтось отримає доступ до ваших таблиць, також не варто. За замовчуванням спільний доступ відключений, а самі файли зберігаються на сервері в зашифрованому вигляді. В Google Spreadsheets можуть бути дані чотирьох типів - це цифри, текст, дати і формули. Формули починаються з знаку «=». В осередках допускається використання URL, а дані в стовпцях можна сортувати. На рис. 3.2 наведений приклад формул у таблицях програми. [79].

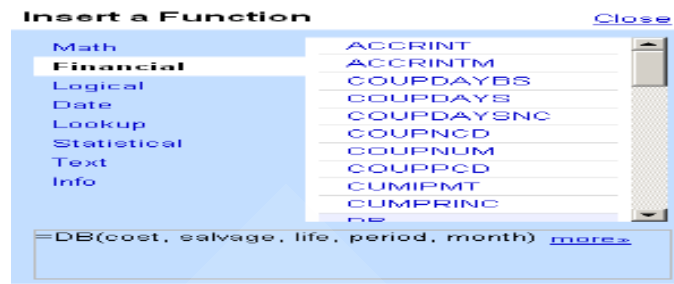


Рисунок 3.2 – Формули у таблицях Google Spreadsheets

У Google Spreadsheets поки досить багато слабких місць. Перш за все, відзначимо неможливість створення і вбудовування графіків і діаграм, не допускається використання макросів. У порівнянні з Excel 2003 не так багато вбудованих функцій (functions).

Але є і вагомні переваги. Зберігання даних онлайн, дозволяє працювати з ними з будь-якого комп'ютера, де встановлений MS IE6 + або Mozilla Firefox 1.0.7+. На особливу увагу заслуговують можливості колективної роботи над документом – користувач може визначити, хто має право редагувати таблицю, а хто тільки переглядати. Спільна робота – найважливіша перевага Google Spreadsheets. Одночасно переглядати або правити таблицю можуть до 50 осіб. Причому передбачено чат, щоб користувачі могли обговорювати та вносити зміни. [79]

Google Spreadsheets спрямована на використання новітніх інформаційно-комунікаційних технологій та програм, що допомагають вдосконалити управління, покращують взаємодію працівників та сприяють соціальним інноваціям.

Google таблиці (Google Spreadsheet) це зручний спосіб для парсинга, якщо немає необхідності парсити велику кількість даних, так як є ліміти на кількість xml запитів в день.

За допомогою таблиць Google Spreadsheet можна парсити метадані, заголовки, найменування товарів, ціни, пошту і багато іншого.

Розглянемо основні функції:

-функція *importHTML*:

налаштовує імпорт таблиць і списків на сторінках сайту.

Прописується наступним чином:

= IMPORTHTML ("посилання на сторінку"; запит "table" або "list"; порядковий номер таблиці / списку)

Наведемо приклад використання:

Необхідно вивантажити дані з таблиці зі сторінки сайту (див. рис. 3.3).



Температура замерзания, °C	-50	-60	-75	-85	-60	-50	-40	-30	-20	-10
Содержание этиленгликоля в %	78,4	72,1	66,7	66	63,1	58	52,6	45,6	36,4	26,4
Плотность этиленгликоля, кг/м.куб.	1088,3	1092,3	1085,6	1084,8	1083,3	1078	1071,3	1062,7	1050,6	1034

Наименование	Описание	Варианты фасовки	Вес нетто, кг	Кол-во единиц в паллете	Цена за кг (с НДС, до 5т)	Цена за кг (с НДС, от 5т до 10т)	Цена за кг (с НДС, от 10т)
Этиленгликоль (МЭГ)		Налие	1	1	90,45 руб. Купить	89,60 руб. Купить	88,70 руб. Купить
Этиленгликоль (МЭГ)	Этиленгликоль высшего сорта ГОСТ 19710-83	Бочка	220	4	103,45 руб. Купить	94,05 руб. Купить	87,50 руб. Купить
Этиленгликоль (МЭГ)		Еврокуб	1113	1	106,75 руб. Купить	97,05 руб. Купить	90,30 руб. Купить

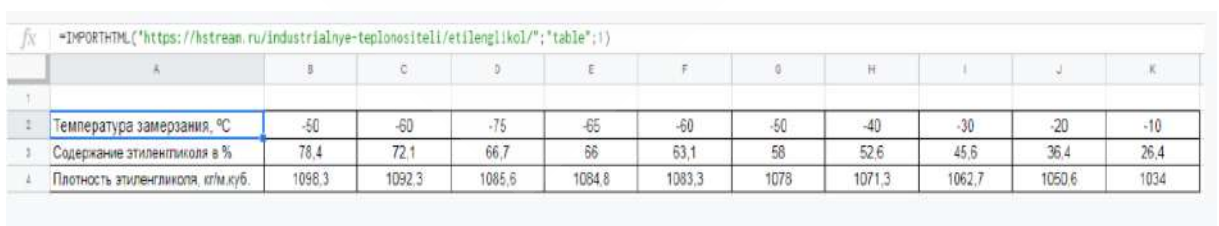
Рисунок 3.3 – Приклад вивантаження даних зі сторінки сайту за допомогою таблиць Google Spreadsheet

Для цього в формулу поміщаємо URL сторінки, додаємо тег «table» і порядковий номер - 1.

Ось що виходить:

= IMPORTHTML ("https://hstream.ru/industrialnye-teplonositeli/etilenglikol/"; "table"; 1)

Вставляємо формулу в таблицю і дивимося результат (див.рис.3.4).



	A	B	C	D	E	F	G	H	I	J	K
1											
2	Температура замерзания, °C	-50	-60	-75	-85	-60	-50	-40	-30	-20	-10
3	Содержание этиленгликоля в %	78,4	72,1	66,7	66	63,1	58	52,6	45,6	36,4	26,4
4	Плотность этиленгликоля, кг/м.куб.	1088,3	1092,3	1085,6	1084,8	1083,3	1078	1071,3	1062,7	1050,6	1034

Рисунок 3.4 – Результат та команда, яку було задано формулі IMPORTHTML при порядковому номері 1

Для вивантаження другий таблиці у формулі замінюємо 1 на 2.

= IMPORTHTML ("https://hstream.ru/industrialnye-teplonositeli/etilenglikol/"; "table"; 2) [78].

Вставляємо формулу в таблицю і дивимося результат (див. рис. 3.5).

	A	B	C	D	E	F	G	H	I
1									
2	Наименование	Описание	Варианты фасов	Вес нетто, кг	Кол-во единиц в	Цена за кг (с НДС)	Цена за кг (с НДС)	Цена за кг (с НДС)	от 10т*
3	Этиленгликоль (МЭГ)	Этиленгликоль	Налив		1	90,45 руб. Купить	89,60 руб. Купить	88,70 руб. Купить	
4	Этиленгликоль (МЭГ)		Бочка	220	4	103,45 руб. Купить	94,05 руб. Купить	87,50 руб. Купить	
5	Этиленгликоль (МЭГ)		Еврокуб	1113	1	106,75 руб. Купить	97,05 руб. Купить	90,30 руб. Купить	
6									

Рисунок 3.5 – Результат та команда, яку було задано формулі IMPORTHTML при порядковому номері 2

-функція importXML:

імпортує дані з документів в форматах HTML, XML, CSV, CSV, TSV, RSS, ATOM XML.

Функція має більш широкий спектр опцій, ніж попередня. З її допомогою зі сторінок і документів можна збирати інформацію практично будь-якого виду.

Робота з цією функцією передбачає використання мови запитів XPath.

Формула:

= IMPORTXML ("посилання"; "// XPath запит")

Наведемо приклад використання:

Витягуємо title, description і заголовок h1.

У першому випадку у формулі прописуємо // title:

= Importxml (A3; "// title")

Вставляємо формулу в таблицю і дивимося результат (див. рис. 3.6).

fx =importxml(A3;"//meta[@name='description']/@content")			
	A	B	C
1			
2	URL	Title	Description
3	https://www.skyline-geo.ru/uslugi/mezhevanie-uchastkov	Межевание земельного участка - цена, заказать межевой план в Санкт-Петербурге и Ленинградской области	Заказать работы по межеванию земельного участка с подготовкой межевого плана по выгодной стоимости в Санкт-Петербурге и Ленинградской области в компании «СкайЛайн».

Рисунок 3.8 – Результат та команда, яку було задано формулі IMPORTXML

У випадку з іншими будь-якими даними XPath можна скопіювати прямо з коду сторінки.

Робиться це як наведено на рис. 3.9.

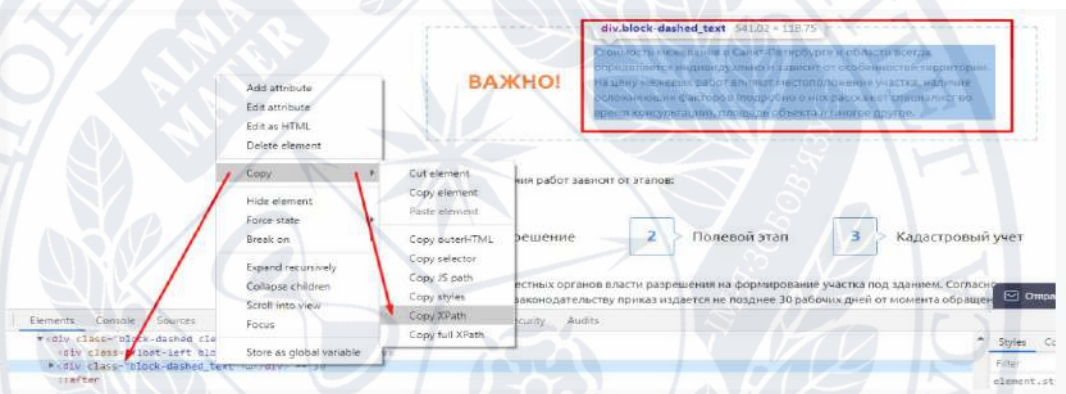


Рисунок 3.9 – Дані XPath прикопіюванні з коду сторінки

Ось як це буде виглядати після всіх маніпуляцій
 = Importxml (A3; "// html / body / div / div / div [1] / div [2] / div / div [4] / div [2]"))

Вставляємо формулу в таблицю і дивимося результат (див. рис. 3.10). [80].

fx		=importxml(A3;"//html/body/div/div/div[1]/div[2]/div/div[4]/div[2]")
	A	B
1		
2	URL	Текст блока
3	https://www.skyline-geo.ru/uslugi/mezhevanie-uchastkov	Стоимость межевания в Санкт-Петербурге и области всегда определяется индивидуально и зависит от особенностей территории. На цену межевых работ влияют местоположение участка, наличие осложняющих факторов (подробно о них расскажет специалист во время консультации), площадь объекта и многое другое.

Рисунок 3.10 – Результат та команда, яку було задано формулі IMPORTXML

-функція **REGEXEXTRACT**:

З її допомогою можна отримувати будь-яку частину тексту, яка відповідає регулярному виразу.

Звичайно для використання даної функції необхідні знання побудови регулярних виразів,

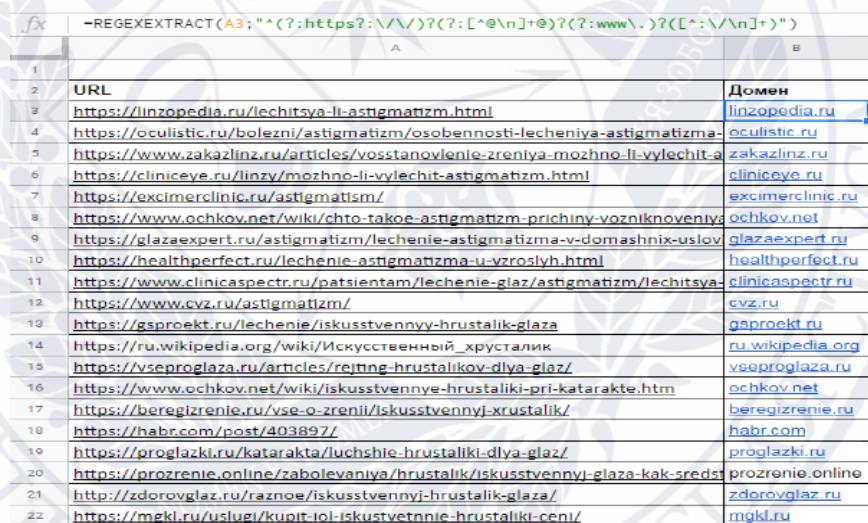
Наведемо приклад використання:

Потрібно відділити домени від сторінок.

Це можна зробити за допомогою формули:

=REGEXEXTRACT (A2; "^ (?: https?: \\ / \\ /)? (?: [^ @ \\ N] + @)? (?: www \\ .)? ([^: \\ / \\ N] +) ")

Вставляємо формулу в таблицю і дивимося результат (див. рис. 3.11).



URL	Домен
https://linzopedia.ru/lechitsya-li-astigmatizm.html	linzopedia.ru
https://oculistic.ru/bolezni/astigmatizm/osobennosti-lecheniya-astigmatizma-	oculistic.ru
https://www.zakazlinz.ru/articles/vosstanovlenie-zreniya-mozhno-li-vylechit-a	zakazlinz.ru
https://cliniceye.ru/linzy/mozhno-li-vylechit-astigmatizm.html	cliniceye.ru
https://excimerclinic.ru/astigmatizm/	excimerclinic.ru
https://www.ochkov.net/wiki/chto-takoe-astigmatizm-prichiny-vozniknoveniya	ochkov.net
https://glazaexpert.ru/astigmatizm/lechenie-astigmatizma-v-domashnix-uslov	glazaexpert.ru
https://healthperfect.ru/lechenie-astigmatizma-u-vzroslyh.html	healthperfect.ru
https://www.clinicaspectr.ru/patsientam/lechenie-glaz/astigmatizm/lechitsya-	clnicaspectr.ru
https://www.cvz.ru/astigmatizm/	cvz.ru
https://gsproekt.ru/lechenie/iskusstvennyy-hrystalik-glaza	gsproekt.ru
https://ru.wikipedia.org/wiki/Искусственный_хрусталик	ru.wikipedia.org
https://vseproglaza.ru/articles/rejting-hrystalikov-dlya-glaz/	vseproglaza.ru
https://www.ochkov.net/wiki/iskusstvennye-hrystaliki-pri-katarakte.htm	ochkov.net
https://beregizrenie.ru/yse-o-zrenii/iskusstvennyj-xrystalik/	beregizrenie.ru
https://habr.com/post/403897/	habr.com
https://proglazki.ru/katarakta/luchshie-hrystaliki-dlya-glaz/	proglazki.ru
https://prozrenie.online/zabolevaniya/hrystalik/iskusstvenny-glaza-kak-sredst	prozrenie.online
http://zdorovglaz.ru/raznoe/iskusstvennyj-hrystalik-glaza/	zdorovglaz.ru
https://mgkl.ru/uslugi/kupit-iol-iskustvennie-hrystaliki-ceni/	mgkl.ru

Рисунок 3.11 – Результат та команда, яку було задано формулі IMPORTXML при додаванні осередків, в яких містяться необхідні дані

Виходячи з прикладів та розбору програми, можна навести переваги та недоліки програми та надати рекомендації, щодо усунення наведених недоліків. [81].

У автоматичного збору інформації купа переваг (порівняно з ручним методом):

- Програма працює самостійно. Не доводиться витрачати час на пошук та сортування даних. До того ж збирає вона інформацію набагато швидше за людину. Та ще й робить це 24/7, якщо знадобиться.
- Парсеру можна «готувати» стільки параметрів, скільки потрібно, і ідеально відбудувати його для пошуку лише контент. Без сміття, помилок і невиправданої інформації з невідповідних сторінок.
- На відміну від людини, парсер не допускати дурних помилок через неухважність. І не втомиється.
- Утиліта для парсингу може подавати знайдені дані у зручному форматі на запит користувача.
- Парсери вміють грамотно розподіляти навантаження на сайт. Це означає, що він випадково не «впустить» чужий ресурс, а вас не звинуватить у незаконній DDoS-атаці. [82].

Підсумовуючи вищесказане, можна зазначити, що програми-парсери істотно прискорюють процес роботи особливо, якщо є ключові слова. Налаштувавши роботу, можливо оперативно підібрати необхідні для просування запити, тому що, за короткий термін обходить тисячі сторінок, фільтрує представлені дані, відбираючи серед них потрібні, після чого пакує отриманий результат для подальшої обробки. Адже, це дає можливість відстежувати зміни подій в закладі освітньої установи, швидко копіювати інформацію з інших сайтів для розміщення на власному ресурсі тощо.

Отже немає сенсу «парсити» руками, коли можна довірити цю операцію відповідному ПЗ.

Але головний недолік парсерів у тому, що не завжди вдається скористатися. Зокрема, коли власники чужих сайтів забороняють автоматичний збір інформації зі сторінок. Є відразу кілька методів блокування доступу з боку парсерів: і IP-адресами, і за допомогою налаштувань для пошукових ботів. Усі вони досить ефективно захищають від парсингу. [83].

У мінуси методу можна віднести і те, що конкуренти також можуть використовувати його. Щоб захистити сайт від парсингу, доведеться вдатися до однієї з технік:

- або заблокувати запити з боку, вказавши відповідні параметри robots.txt;
- або налаштувати капчу - навчити парсер розгадуванню картинок занадто затратно, ніхто цим не займатиметься.

Але всі методи захисту легко обходяться, тому швидше за все доведеться з цим явищем миритися.

Отже, не завжди зрозуміло, як визначати сам термін "пошукова машина", тому що навіть спайдери каталогів інколи шукають інформацію і повертають результати. Зрозуміло, в пошукових системах немає редакторів, що переглядають мільярди сторінок, аби особисто перевірити їх якість. І каталоги теж не можуть дати всім користувачам саме те, що вони шукають. Таким чином, для пошукових систем доцільніше розставити певні пріоритети перевірених редакторами надійних каталогів сторінок. Пошукові технології, використовувані краулерами, придатні і для каталогів що допомагає долати обмеження маленьких баз даних. [84].

Висновки до розділу 3

Отже, парсинг сайтів є ефективним рішенням для автоматизації збору і зміни інформації. У порівнянні з людиною, програма-парсер здійснює 4 кроки при опрацюванні масиву даних: *швидко обійде тисячі веб-сторінок; обережно відокремить технічну інформацію від «людської»; безпомилково відбере потрібне і відкине зайве; ефективно упакує кінцеві дані в необхідному вигляді.*

Наведено конкретний приклад програми-парсеру. Це Google Spreadsheet - програма електронних таблиць, який використовуватиме відділ кадрів Донецького національного університету імені Василя Стуса.

Google Spreadsheet –

Наведено характеристики даної програми, яка успішно допомагатиме у пошуку тої чи іншої потрібної інформації:

1. Інтерфейс програми-парсеру Google Spreadsheets.
2. Імпорт та експорт даних та файлів.
3. Робота Google Spreadsheets.

У Google Spreadsheets поки досить багато слабких місць. Перш за все, відзначено неможливість створення і вбудовування графіків і діаграм, не допускається використання макросів.

Але є і вагомі переваги. Зберігання даних онлайн, дозволяє працювати з ними з будь-якого комп'ютера, де встановлений MS IE6 + або Mozilla Firefox 1.0.7+.

Google Spreadsheets спрямована на використання новітніх інформаційно-комунікаційних технологій та програм, що допомагають вдосконалити управління, покращують взаємодію працівників та сприяють соціальним інноваціям. [42]

За допомогою таблиць Google Spreadsheet можна парсити метадані, заголовки, найменування товарів, ціни, пошту і багато іншого.

Було розглянути такі основні функції: **-функція *importHTML***; **-функція *importXML***; **-функція *REGEXEXTRACT***.

Виходячи з прикладів та розбору програми, було наведено переваги та недоліки програми та надано рекомендації, щодо усунення наведених недоліків. [84].

ВИСНОВКИ

Розкрито сутність «парсинг», «програми-парсери» на основі нормативно-правової бази та наукових здобутків. Парсер – програма, що забезпечує парсинг, тобто синтаксичний аналіз контенту в мережі до певної математичної моделі, створеної на одній з мов програмування, зазвичай на Python, PHP або Java.

Висвітлено законодавчу та нормативно-правову базу діяльності парсерів. Оскільки парсинг – це один із способів збору інформації, то цей процес підпорядковується правилам, встановленим для поводження з інформацією, зокрема, Закону України «Про інформацію».

У свою чергу інформація з обмеженим доступом поділяється на конфіденційну, таємну та службову. Визначення кожного виду містяться в Законі України «Про доступ до публічної інформації».

Проаналізовано використання програм-парсерів у роботі на прикладі компанії. Наприклад, компанія AVADA MEDIA пропонує як готові рішення в даній області, так і парсери на замовлення

Проаналізовано комунікаційний процес закладу вищої освіти. В даному випадку комунікація Донецького національного університету імені Василя Стуса відзначається такими рисами, як: *розосередженість аудиторії; наявність широкомасштабних комунікаційних процесів; одночасна передача даних різними каналами масового інформування тощо.*

Визначено передумови впровадження програми-парсеру у відділі кадрів: є обробка великого масиву інформації в найкоротші терміни; у звіті не буде помилок; ці програми не тільки будуть збирати дані, а й будуть рекомендувати, що потрібно видалити, та які помилки виправити тощо.

Проведено технічний і інноваційний аналіз функціонування програми-парсеру у відділі кадрів та визначено проблеми впровадження методу парсингу в закладах вищої освіти в напрямі забезпечення інформацією відділу кадрів, як небажання вивчати нову програму, юридична обґрунтованість облікових даних та контроль за правами доступу.

Визначено перспективи використання програм-парсерів у відділі кадрів для удосконалення інформаційної діяльності закладів вищої освіти: програми-парсери спрямовані на використання новітніх інформаційно-комунікаційних технологій та програм, що допомагають вдосконалити управління, покращують взаємодію працівників та сприяють соціальним інноваціям.



СПИСОК ВИКОРИСТАНИХ ПОСИЛАНЬ

1. Маркетингове агентство МАВР. Особистий блог Владислави Рикової. Datasol – Парсинг для збору інформації з сайтів. URL: <https://vladykova.com/parser-sajtov/> (дата звернення: 22.08.2021).
2. Компанія-розробник ІТ-продуктів AVADA MEDIA™. Парсери, що це і де вони використовуються. URL: <https://avada-media.ua/ua/services/parser/> (дата звернення: 23.08.2021)
3. Компанія-розробник ІТ-продуктів AVADA MEDIA™. Парсери. Що таке програми-парсери і як вони працюють. URL: <https://avada-media.ua/ua/services/parser/> (дата звернення: 30.08.2021)
4. Компанія-розробник ІТ-продуктів AVADA MEDIA™. Парсери. Навіщо вони і чи потрібні. URL: <https://avada-media.ua/ua/services/parser/> (дата звернення: 30.08.2021)
5. Geitgevey A. Natural Language Processing is Fun! URL: <https://medium.com/@ageitgey/natural-language-processing-is-fun-9a0bff37854e> (дата звернення: 01.09.2021)
6. Парсинг сайтів. Що це і як працює. URL: <https://www.business-gazeta.ru/article/453414> (дата звернення: 01.09.2021)
7. Кащенко Д.О. Дослідження характеристик веб-застосунків ВНЗ за допомогою робота-парсера: автореф. маг. н. р. на здобуття освіт. кваліфікації «Маг. сист. ан.»: 004.89. Миколаїв, 2020. 83 с.
8. Ханенко О.А. Автоматизовані засоби парсингу проектів дистанційного навчання в Cloud для заповнення сайту <http://dl-cloud.kpi.ua>: дипломна. Київ: КПІ, 2016. 75 с.
9. Що таке парсинг і коли його використовувати? URL: <https://alexsmokinof.lviv.ua> (дата звернення: 01.09.2021)
10. World of science and technologies. Про науку та технології в Україні та світі. Що таке пошукові машини і як вони працюють? URL:

<https://web.archive.org/web/20151222170219/http://scitechspace.blogspot.com/2014/09/blog-post.html> (дата звернення: 01.09.2021)

11. Копірайтинг-агенція Weblana. Парсинг сайтів – плюси та мінуси використання. Що таке парсити сайт. Навіщо потрібна програма парсингу сайту. URL: <https://weblana.com/a403615-parsing-sajtov-plyusy.html> (дата звернення: 15.09.2021)

12. Небезпеки парсингу. Чим може обернутися парсинг даних із сайту. URL: <https://weblana.com/a403615-parsing-sajtov-plyusy.html> (дата звернення: 01.09.2021)

13. Парсинг. Різновиди парсингу. Етапи парсингу. URL: <https://www.seonews.ru/glossary/parsing>. (дата звернення: 15.09.2021)

14. Парсинг. Парсери. Використання парсерів. Алгоритми роботи парсерів. Основи роботи парсерів. Етичні і технічні складнощі роботи з парсерами. URL: <https://ipipe.ru/info/parsing>. (дата звернення: 16.09.2021)

15. Висоцька В.А. Особливості моделювання синтаксису речення слов'янських та германських мов за допомогою породжуваних контекстно-вільних граматик. Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі. 2015. 246-276 с.

16. Історія синтаксичного аналізу (парсерів). URL: <https://tproger.ru/translations/parsing-a-timeline/> (дата звернення: 20.09.2021)

17. Парсинг: від теорії до судової практики. Правове регулювання парсингу. URL: <https://lcf.ua/news/pravove-reguluvannia-parsyngu/> (дата звернення: 16.09.2021)

18. Що таке парсинг. Обмеження при програмі-парсеру. Законність парсерів. Як захиститися від парсерів. URL: <https://blog.calltouch.ru/chto-takoe-parsing>. (дата звернення: 20.09.2021)

19. Парсери. Парсинг. Використання програм-парсерів. URL: <https://alexsmokinof.lviv.ua>. (дата звернення: 20.09.2021)

20. Синтаксичний аналіз. Мови програмування. Парсери. Програми-парсери. URL: <http://xn--r1a3b.xn--b1amgblet.xn--j1amh/index.ph>. (дата звернення: 25.09.2021)

21. Парсинг: від теорії до практики. Правове регулювання парсингу (парсерів). URL: https://jurliga.ligazakon.net/ua/analitycs/194822_parsing-vd-teor-do-sudovo-praktiki. (дата звернення: 25.09.2021)

22. Вакуленко Ю.О. Роль прес-кліпінгу в рекламній діяльності підприємства / Ю.О. Вакуленко, О. С. Щербіна // Вісник студентського наукового товариства Донецького національного університету імені Василя Стуса. – Вінниця: ДонНУ імені Василя Стуса, 2018. – Випуск 10. – Том 2. – С. 132-135.

23. Колісниченко Д.М. Пошукові системи і просування сайтів. Діалектика, 2014. 272 с.

24. Семантика. Парсинг сайту. Етапи парсингу даних. URL: semantica.in/blog/chto-takoe-parsing. (дата звернення: 30.09.2021)

25. Frost, R., Hafiz, R. and Callaghan, P. (2008) « Parser Combinators for Ambiguous Left-Recursive Grammars.» *10th International Symposium on Practical Aspects of Declarative Languages (PADL), ACM-SIGPLAN*, Volume 4902/2008, Pages:, January 2008, San Francisco. 167-181 pp.

26. Про інформацію: Закон України від 02.10.1992. №48. 2021, С.20-50.

27. Про доступ до публічної інформації: Закон України від 13.01.2011. №32. 2021, С.15-45.

28. Гораш І. Веб-сайт як популярне інформаційне джерело / І. Гораш, Г. П. Лукаш // Інформаційні технології і системи в документознавчій сфері: IV Всеукраїнська наукова студентська конференція, м. Вінниця, 04 квітня 2019 р. – Вінниця: ДонНУ імені Василя Стуса, 2019. 28-30 с.

29. Терещенко В. В. Перспективи вдосконалення інформаційного пошуку. URL: <http://molodyvcheny.in.ua/files/journal/2017/4.4/25.pdf>. (дата звернення: 30.09.2021)

30. Ашманов І.С. Просування сайту в пошукових системах. Вільямс, 2016. 304 с.

31. Менеджмент. Організаційна поведінка. Організаційні комунікації в установах. URL: https://pidru4niki.com/menedzhment/organizatsiyni_komunikatsiyi. (дата звернення: 30.09.2021)

32. Вакуленко Ю. Використання інформаційних систем і технологій у закладах вищої освіти Вінниці / Ю. Вакуленко, О. С. Щербіна // Інформаційні технології і системи в документознавчій сфері: III Всеукраїнська наукова студентська конференція, м. Вінниця, 05 квітня 2018 р. – Вінниця: ДонНУ імені Василя Стуса, 2018. 16-18 с.

33. Парсинг html-сайтів для споживача. Парсинг даних для використання підприємством. URL: parsing.valemak.com/ru/what-why-how/stages-of-parsing. (дата звернення: 01.10.2021)

34. «Баст» веб-студія. Парсинг сайту. Послуги парсинга сайтів. Переваги парсингу. URL: webstudiobast.com/uk/poslugu-parsinga-sajtiv-dlya-opencart. (дата звернення: 01.10.2021)

35. Парсинг. Метод парсингу. Де використовується парсинг. Особливості його застосування. URL: alexsmokinof.lviv.ua. (дата звернення: 01.10.2021)

36. Парсинг. URL: <http://parsing.valemak.com>. (дата звернення: 02.10.2021)

37. Методи парсингу сайтів. URL: <http://ponka.vnukov.ru/content/metody-parsinga-saytov>. (дата звернення: 02.10.2021)

38. Щербіна О.С., Вакуленко Ю.О. Впровадження та перспективи використання програм-парсерів у закладах вищої освіти. *Бібліотекознавство. Документознавство. Інформологія: наук. журн. / М-во культури та інформ. політики України, Нац. акад. керів. кадрів культури і мистецтв. Київ: НАКККиМ, 2021. №2. 88-95 с.*

39. Гораш І. Комунікативний потенціал управлінської діяльності / І. Гораш, Г. П. Лукаш // Інформаційні технології і системи в документознавчій

сфері: V Всеукраїнська наукова студентська конференція, м. Вінниця, 15 квітня 2020 р. – Вінниця: ДонНУ імені Василя Стуса, 2020. 6-8 с.

40. Грищенко І., Крахмальова Н. Роль комунікацій у функціонуванні вищих навчальних закладів. URL: https://knutd.edu.ua/publications/pdf/International_editions/Gryshchenko_Krakhmal. (дата звернення: 02.10.2021)

41. Менеджмент. Комунікації в менеджменті: сутність та їх види. URL: <https://ru.osvita.ua/vnz/reports/management/14536>. (дата звернення: 02.10.2021)

42. Експерт з якості освіти. Донецький національний університет імені Василя Стуса. URL: <https://www.donnu.edu.ua/uk/ekspertu-z-yakosti-vyshhoi-osvity>. (дата звернення: 02.10.2021)

43. Інформація, комунікація та управління знаннями в глобалізаційному світі. URL: <http://knukim.edu.ua/wp-content/uploads/2020/09/information-komunikation> (дата звернення: 03.10.2021)

44. Маркетингові дослідження. Система інтегрованих маркетингових комунікацій. URL: https://pidru4niki.com/marketing/sistema_komunikatsiy. (дата звернення: 03.10.2021)

45. Документно-інформаційні комунікації в умовах глобалізації: стан, проблеми, перспективи. URL: <http://dspace.wunu.edu.ua/bitstream>. (дата звернення: 03.10.2021)

46. Менеджмент. Комунікації в організаціях: процес, типи, зв'язки, бар'єри. URL: <https://ru.osvita.ua/vnz/reports/management/15040>. (дата звернення: 03.10.2021)

47. Меннінг К. Введення в інформаційний пошук. Вільямс, 2011. 600 с.

48. Brin S., Page L. The anatomy of a large-scale hypertextual Web search engine // Computer Networks and ISDN Systems, 2004. 107-117 pp.

49. Щербіна О.С. Вакуленко Ю.О. Концепція формування комунікації з використанням програм-парсерів у відділі кадрів закладу вищої освіти. *Інформаційні технології і системи в документознавчій сфері*: VI Всеукраїнська

наукова студентська конференція, м. Вінниця, 02 квітня 2021 р. Вінниця: ДонНУ, 2021. 17-19 с.

50. Маркетинг. Маркетингові комунікації. Внутрішньокорпоративні комунікації. URL: https://stud.com.ua/67208/marketing/vnutrishnokorporativnih_komunikatsiy (дата звернення: 04.10.2021)

51. Маркетинг. Зв'язки з громадськістю. Корпоративні комунікації. URL: https://stud.com.ua/143013/marketing/korporativni_komunikatsiyi (дата звернення: 04.10.2021)

52. Територія розвитку та освіти Оксани Королович. Види комунікації в організації. Психологія бізнесу і організацій. URL: <http://oksanakorolovych.com/uk/vidi-komunikatsii-v-organizatsii> (дата звернення: 05.10.2021)

53. Маркетинг. Реклама і зв'язки з громадськістю. Засоби внутрішньоорганізаційних комунікацій. Канали розповсюдження інформації всередині фірми. URL: https://stud.com.ua/18337/marketing/zasobi_vnutrishnoorganizatsiyni_komunikatsiy_kanali_rozpovsyudzhennya_informatsiyi_vseredini_firmi (дата звернення: 05.10.2021)

54. Менеджмент. Види комунікацій. URL: https://pidru4niki.com/13731120/menedzhment/vidi_komunikatsiy (дата звернення: 05.10.2021)

55. Внутрішня комунікація: як ефективно побудувати діалог в колективі. URL: <https://gurt.org.ua/articles/39901/> (дата звернення: 05.10.2021)

56. Комунікації в організаціях. URL: <https://cutt.ly/HTC0tTH> (дата звернення: 05.10.2021)

57. Костенко П.П., Левченко І.В. Веб-сервіс уточнення релевантності веб-документів пошукової видачі Google на основі поведінки користувача // Інженерні та освітні технології. Щоквартальний науково-практичний журнал – Кременчук: КрНУ, 2014. – Випуск 4(8). 49-62 с.

58. Вакуленко Ю.О. Використання методу парсингу для підвищення ефективності діяльності підприємства у конкурентному середовищі / Ю.О. Вакуленко, О. С. Щербіна // Інформаційні технології і системи в документознавчій сфері: IV Всеукраїнська наукова студентська конференція, м. Вінниця, 04 квітня 2019 р. – Вінниця: ДонНУ імені Василя Стуса, 2019. 17-19 с.

59. Менеджмент. Методи прийняття управлінських рішень. Основні види комунікацій. URL: https://stud.com.ua/45783/menedzhment/osnovni_vidi_komunikatsiy (дата звернення: 07.10.2021)

60. KMBS. Внутрішні комунікації і корпоративна культура. Стратегічний рівень. URL: <https://kmbs.ua/ua/edp/strategy/vnutrishni-komunikatsiyi-i-korporativna-kultura-strategichniy-riven> (дата звернення: 07.10.2021)

61. Агробізнес Сьогодні. Внутрішні комунікації: вчимося розмовляти. URL: <http://agro-business.com.ua/agro/kermo-kerivnyka/item/1989-vnutrishni-komunikatsii-vchymos-rozmovliaty.html> (дата звернення: 08.10.2021)

62. Хайлайти вебінару Яни Бакун. 10 порад для покращення внутрішніх комунікацій. URL: <https://dtek.com/media-center/news/10-sovetov-dlya-uluchsheniya-vnutrennikh-kommunikatsiy/> (дата звернення: 08.10.2021)

63. Together Європейський Простір. Комунікація всередині організації. URL: <https://euprostor.org.ua/courses/lecture/149142> (дата звернення: 15.10.2021)

64. Система зовнішньої і внутрішньої комунікації в діяльності підприємства. URL: <https://cutt.ly/9TC3wG1> (дата звернення: 16.10.2021)

65. Маркетинг. Ділові комунікації. Засоби внутрішньоорганізаційних комунікацій. URL: https://studme.com.ua/1566072111130/marketing/sredstva_vnutriorganizatsionnyh_kommunikatsiy.htm (дата звернення: 18.10.2021)

66. Менеджмент. Комунікації в організаціях: процес, типи, зв'язки, бар'єри.

URL: <https://ru.osvita.ua/vnz/reports/management/15040/> (дата звернення: 20.10.2021)

67. Пошукова система. Функції, механізми програм. URL: https://www.spip.net/uk_article5780.html (дата звернення: 25.10.2021)

68. Технічна енциклопедія Tech Trend. Пошукова машина, механізми функціонування. URL: <http://techtrend.com.ua/index.php?newsid=14798> (дата звернення: 25.10.2021)

69. Пошукові системи: склад, функції, механізми, принципи роботи. URL: <https://shag.com.ua/poshukovi-sistemi-sklad-funkciyi-princip-roboti.html> (дата звернення: 25.10.2021)

70. Соціальна інформатика. Інформаційно-пошукові системи (ІПС). URL: <https://sites.google.com/site/socialnafnformatika/home/informacijno-posukovi-sistemi> (дата звернення: 26.10.2021)

71. Інформатика та комп'ютерна техніка. Мережеві технології. Пошук, обмін та захист даних і Інтернеті. URL: <https://kppk.com.ua/ELLIB/ebook/Gorbenko/IKT/14/14.htm> (дата звернення: 26.10.2021)

72. Що таке пошукові системи? Функціонування пошукових систем. URL: <https://kovelpost.com/blogs/103> (дата звернення: 26.10.2021)

73. Техніка. Технології. Принцип роботи пошукової машини. URL: <https://piznayko.in.ua/pryntsypu-roboty-poshukovoyi-mashyny-rambler> (дата звернення: 27.10.2021)

74. Говорущенко Т. О. Павлова О. О. Боднар М. А. Сучасні проблеми семантичного аналізу специфікацій вимог до програмного забезпечення. Вчені записки ТНУ імені В.І. Вернадського. Серія: Технічні науки. URL: <http://tech.vernadskyjournals.in.ua/journals/2019>. (дата звернення: 28.10.2021)

75. Пошукові системи та механізми. Принцип роботи пошукових систем. URL: <https://sites.google.com/site/srchngns/family-profiles> (дата звернення: 28.10.2021)

76. Структура та принципи функціонування пошукової системи у ВЕБ.
URL: <http://www.compiko.lviv.ua/wp-content/uploads/tip/tip10.pdf> (дата звернення: 28.10.2021)
77. Спеціалізовані пошукові системи для глобальної мережі. URL: <http://inmad.vntu.edu.ua/svbevz/p138.pdf> (дата звернення: 28.10.2021)
78. Grune D. Parsing Techniques - A Practical Guide / D. Grune, C. Jacobs.
– Chichester: Originally published by Ellis Horwood, 1990. 320 p.
79. Popular Technology. URL: http://www.populartechology.net__ (дата звернення: 29.10.2021)
80. Google Spreadsheets. Google Spreadsheets – онлайнний табличний процесор. Аналітика. Переваги Google Spreadsheets. URL: internetno.net./google-spreadsheets-onlaynovyi-y-tablichnyi-y-protssessor. (дата звернення: 29.10.2021)
81. Базові поради для роботи з Google Spreadsheets. Google Spreadsheets та таблиці. URL: blog.ringostat.com/ru/bazovie-sovety-dlya-udobnoy-raboty-v-google-spreadsheets. (дата звернення: 30.10.2021)
82. Візуалізації в Google Spreadsheets. Візуалізації в таблицях. Google Spreadsheets. URL: habr.com./post/461555. (дата звернення: 30.10.2021)
83. Горах І. Аналіз популярних корпоративних інформаційних систем / І. Горах, Т. В. Січко // Комп'ютерні технології обробки даних: Всеукраїнська науково-практична конференція для студентів, аспірантів та молодих вчених, м. Вінниця, 04 грудня 2020 р. – Вінниця: ДонНУ імені Василя Стуса, 2020. 26-30 с.
84. Трансформаційні процеси у суспільній та соціокультурній сферах України: монографія / авт. колектив: Анісімова О.М., Ковальська Л.А., Лукаш Г.П., Прігунов О.В., Щербіна О.С., Яворська Т.М. Вінниця: ДонНУ імені Василя Стуса, 2021. 176 с.