International Journal of Advances in Applied Sciences (IJAAS)
Vol. 13, No. 4, December 2024, pp. 860~877
ISSN: 2252-8814, DOI: 10.11591/ijaas.v13.i4.pp860-877

The study of requirements for the workforce of the digital industries using web scraping techniques

Rachawit Tipsena[1], Wirapong Chansanam[1], Kulthida Tuamsuk[1], Tossapon Boongoen[2]
[1]Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Khon Kaen, Thailand
[2]Department of Computer Science, Faculty of Business and Physical Sciences, Aberystwyth University, Aberystwyth, United Kingdom

Article Info

Article history:

Received Mar 3, 2024
Revised Jun 20, 2024
Accepted Jul 3, 2024

Keywords:

Digital industries
Natural language processing
Requirements
Web scraping
Workforce

ABSTRACT

This study investigates workforce requirements in Thailand's digital industries, focusing on qualification requirements across five industry groups: hardware and smart devices, software and software services, digital service, digital content, and telecommunication. Employing Python-based web scraping from selected job websites during 2022–2023, the data undergoes natural language processing (NLP) for analysis. Within Group 1 (hardware and smart devices), electrical engineers dominate with 92 positions, emphasizing a demand for engineering expertise. Group 2 (software and software services) sees a surge in programmer roles, totaling 244 positions, showcasing a need for robust programming skills. Group 3 (digital service) prioritizes information technology (IT) support, claiming 354 positions, indicating high demand for IT support qualifications. Graphic design leads Group 4 (digital content) with 587 positions, highlighting the need for a workforce in digital content production. In Group 5 (telecommunication), network engineers dominate with 37 positions, signaling a demand for top-tier network engineering skills. Most positions across groups specify a bachelor's degree and often require prior experience, highlighting the industry's preference for both academic and practical qualifications. This study underscores the digital industry's rapid growth and the sustained demand for a qualified workforce, emphasizing the importance of academic credentials and specialized skills for future employment in these sectors.

Corresponding Author:

Wirapong Chansanam
Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University
123 Mitraparp Road, Mueang Khon Kaen, Khon Kaen 40002, Thailand
Email: wirach@kku.ac.th

1. INTRODUCTION

The digital industry, often referred to as industry 4.0, encompasses businesses engaged in the production, distribution, and provision of services related to digital technology, leveraging its advancements. This sector not only plays a pivotal role in propelling the economy but also contributes significantly to the evolution of a digital society [1]. Crucially, the digital industry holds a strategic position in the government's vision for Thailand's economic development and is identified as one of the 10 target industries (S-Curve). The Ministry of Industry has specifically earmarked the digital industry as a key player in steering the nation towards a digital economy, anticipating substantial shifts in investment dynamics [2]. Furthermore, it actively aligns with the government's initiatives in formulating policies for digital labor, aiming to usher in transformative changes in the labor system [3]. However, the advent of digital transformation poses a substantial hiring challenge. This stems from the transformative nature of change, which may usher in new

roles to replace old ones or render some jobs obsolete. Work dynamics that once followed familiar patterns must adapt to change as operational modes become fortified by technology. Notably, the encroachment of machines into tasks once performed by humans is increasingly evident, creating a surge in demand for labor equipped with new skills across various industries [4].

In the digital industry, this demand for skilled labor is particularly pronounced. Entrepreneurs recognize the paramount importance of aligning workers' skills with organizational needs. Employees are expected not only to possess skills that correspond to their job roles but also to continually enhance and adapt their skill sets. This is especially pertinent in the digital industry, where positions such as programmers, developers, system administrators, information technology (IT) system analysts, website designers, and developers remain consistently high in demand [3]. The evolving landscape of the digital industry underscores the necessity for a workforce equipped with dynamic and relevant skills.

Despite the escalating trend of technology replacing human labor, the digital industry continues to present numerous job opportunities. The abundance of job postings on recruitment and general advertisement websites, with over 10 such websites in Thailand alone, showcasing tens of thousands of related job positions, serves as evidence of this [5]. This abundance suggests a sustained and growing demand for workers in the digital industry. Furthermore, job websites serve as easy-to-access platforms for job seekers to explore potential opportunities. These web pages provide detailed information that workers can use to make informed decisions about applying for positions. Simultaneously, these platforms offer employers an effective means of leveraging digital media to disseminate recruitment information. The convenience of online systems also facilitates the streamlined assessment of labor qualifications, eliminating the need for physical visits to apply for positions an efficiency that contributes to cost savings for both employers and job seekers [6]. Even though job recruitment websites specify job titles and the skills or requirements for each position, workers may not be able to determine which job positions and skill sets are most in demand in the digital industry.

Web scraping is the process of using automated scripts to extract data from websites. This method is often used to gather publicly available information but not easily downloadable in a structured format, such as pricing data, contact details, or textual content from web pages [7]. The research by Lunn *et al.* [8] utilizes web scraping and natural language processing (NLP) to study the clear necessities for job applicants, highlighting the in-demand programming skills. This demonstrates that web scraping is beneficial for drawing inferences from large amounts of data. Furthermore, we can employ NLP to reliably extract key information from text. The study by Schedlbauer *et al.* [9] examined the labor market in medical informatics using web scraping to gather data. It highlighted the most frequently found terms, such as programming, experience, projects, and servers. Additionally, it found that 45% of the vocabulary indicates a demand for expert skills, while 55% reflects a demand for soft skills. Research such as that conducted by Priya *et al.* [10] has utilized web scraping to extract data from job portals, enabling a detailed analysis of the job market and the most sought-after skills, as well as those required for specific roles. This information can assist job seekers in better understanding the job market and the skills required to increase their employability. Employers are seeking graduates with skills in emotional intelligence, communication, IT/computer proficiency, problem-solving, analytical abilities, and teamwork capabilities, according to research by MacKenzie [11], which analyzed job descriptions gathered through web scraping. Web scraping also collects job postings from multiple websites and consolidates them on a single platform, thereby enhancing the browsing experience for visitors. This helps reduce the time and complexity involved in searching for information [12]. Additionally, web scraping is used to extract real-time job vacancy data, ensuring that job seekers do not miss out on interesting positions, and it can also suggest skills that workers could improve to enhance their employability [13]. The research also shows how developing potential job search engines using web scraping tools and machine learning techniques can streamline the process of finding jobs that align with job seekers' interests [14].

The burgeoning digital industry is creating a myriad of opportunities and attracting significant investments, resulting in a substantial demand for related job positions. However, this surge in demand necessitates a workforce with specific skills, knowledge (hard skills), abilities, and competencies tailored to each position's requirements. Moreover, while the trend toward digital workplaces presents numerous opportunities for both employers and employees, it also poses clear challenges that require collaborative efforts to overcome. Business leaders will face intense competition when hiring and retaining the best talent. Keeping track of labor market trends related to technology and digital skills will be crucial [15]. As a response to this dynamic landscape, studying related research revealed a research gap in the analysis of the digital industry across the five main industrial sectors, as well as a deficiency in the presentation of job titles that align with the demands of each sector. Thus, this research aims to present comprehensive job requirements and qualifications for individuals entering the labor market in the digital industry, both in the current and future contexts. We collect data from job advertisement websites using web scraping techniques, then extract and process the information using NLP methods to obtain the desired knowledge. The aim is

twofold: first, to provide a guiding framework for those pursuing education in relevant fields of study, enabling them to align their skills with industry needs. Second, it serves as crucial information for existing and prospective workers, aiding them in understanding and developing competencies that harmonize with the evolving operations of the digital industry. This research offers a roadmap for skill development and seeks to contribute to the ongoing enhancement of the workforce in this dynamic and pivotal sector.

2. RELATED WORKS

In this section, the study delves into pertinent literature addressing two focal points central to the core objectives of this research: firstly, the dynamic and ever-evolving landscape of the digital industry, highlighting key trends, challenges, and opportunities that shape its development. Secondly, it examines the increasingly critical role of web scraping as a strategic tool for data collection and analysis within this sector, emphasizing its effectiveness in gathering vast amounts of real-time data. The literature review underscores the importance of adapting to technological advancements and the necessity for innovative methodologies in navigating the complexities of the digital industry. Furthermore, it explores how web scraping contributes to more accurate and comprehensive insights, ultimately informing better decision-making processes.

2.1. The digital industry

This segment scrutinizes existing literature on the digital industry, recognizing it as a pivotal force shaping contemporary life within the digital economy era. It explores the profound impact of emerging technologies on the digital market, highlighting the dynamic nature of digital industrialization that evolves in tandem with technological advancements. The digital industry, encompassing both manufacturing and service sectors, is intricately linked to digital technology utilization. The study categorizes this utilization into five principal groups [16], elucidating the multifaceted dimensions of its influence.

2.1.1. Hardware and smart devices

Focused on the publication and distribution of computer devices, this industry includes smart devices, computer memory, office devices, and implements for the internet of things (IoT), as well as robotics devices. This sector is critical as it drives technological advancements and innovation across various applications, from consumer electronics to industrial automation. The proliferation of IoT and smart technologies is particularly transformative, enhancing connectivity and efficiency in both everyday life and complex industrial systems. As such, the development and distribution of these devices play a pivotal role in shaping future technological landscapes and enabling smarter, more connected environments.

2.1.2. Software and software services

This sector revolves around the creation and dissemination of software products and services. Examples include computer software, data processing software, device control software, and office application software. It serves as the backbone of modern technology infrastructure, providing essential tools for businesses, governments, and consumers to operate efficiently in a digital world. The impact of this sector is widespread, influencing productivity, communication, and data management across all industries. As technology continues to advance, the role of software products and services becomes increasingly vital in enabling innovation and supporting complex systems and processes.

2.1.3. Digital service

This industry provides digital advisory services on digital platforms, such as cloud computing, system maintenance, and IT support. It plays a crucial role in enabling businesses to leverage advanced technologies to improve their operational efficiency, scalability, and security. The focus on cloud computing allows organizations of all sizes to access powerful computing resources and data storage solutions without significant capital investment. Additionally, system maintenance and IT support ensure that these digital systems function optimally, addressing technical issues and preventing downtime. This sector is essential for supporting businesses' digital transformation and helping them adapt to the rapidly changing technological landscape.

2.1.4. Digital content

This sector is engaged in the production, distribution, and service of digital content, which includes computer graphics, animation, digital games, e-books, user interface (UX) or user experience (UI) design, and new media in augmented reality (AR) or virtual reality (VR) formats. This sector is at the forefront of entertainment, education, and marketing innovations, blending creativity with technology to create immersive and engaging experiences for users. Advanced technologies like AR and VR are revolutionizing content

consumption, providing more interactive and realistic experiences that significantly boost user engagement. This dynamic field not only caters to entertainment but also extends its influence to educational platforms and professional training, making it a pivotal area of growth in the digital economy.

2.1.5. Telecommunication

This industry involves the operation of telecommunications network businesses and services, including cellphone network providers, satellite contract providers, and Internet network service providers. It forms the backbone of global communication, facilitating the seamless flow of information across vast distances. By providing critical infrastructure for voice, data, and video transmission, this sector enables both individuals and businesses to connect and communicate effectively, regardless of location. The importance of telecommunications networks has grown exponentially with the digital revolution, supporting everything from everyday mobile communications to complex, cloud-based business solutions. As technology continues to advance, the role of these networks in enabling new forms of digital interaction, such as IoT applications and smart cities, becomes increasingly crucial.

2.2. Web scraping

A data science technique for extracting information from the internet, web scraping, can be executed using packaged software or directly through a programming language. Python, renowned for its efficacy in handling and analyzing large datasets, emerges as the most prevalent programming language for this purpose and is the chosen tool for data collection in this study. The BeautifulSoup library, equipped with user-friendly Python scripts for web scraping, facilitates a streamlined process. This approach allows researchers to efficiently gather and process web data, which is crucial for various analytical applications, thereby enhancing the robustness and depth of research findings.

Python's simplicity allows for concise code creation in just a few steps, making it ideal for data collection and analysis. The process involves importing the necessary library, rendering the web page, establishing storage variables, and extracting and cleaning the data. Once completed, the collected information is stored in a preferred format; in this research, the CSV and Excel file formats are the chosen mediums [7]. This approach ensures efficient and accessible data collection and analysis, aligning perfectly with the study's objectives. Python's ability to work seamlessly with data, coupled with libraries like Pandas and openpyxl for handling CSV and Excel files, respectively, greatly enhances the productivity and effectiveness of research workflows.

3. METHODS AND MATERIALS

The primary unit of analysis for this study comprises data sourced from job posting and advertisement websites, offering a crucial insight into evolving employment trends and skill demands. The aim is to examine the future workforce requirements in the digital industry, classified into five distinct groups: i) hardware and smart devices, ii) software and software services, iii) digital service, iv) digital content, and v) telecommunication. Contemporary technology, particularly information and communications technology, intricately links each of these industries, reflecting broader economic, and technological transformations. Notably, there is a discernible trend indicating an anticipated increase in demand for workforce requirements in the future, underscoring the urgency for educational institutions and policymakers to adapt and respond to these emerging needs. Figure 1 summarizes the workflow.

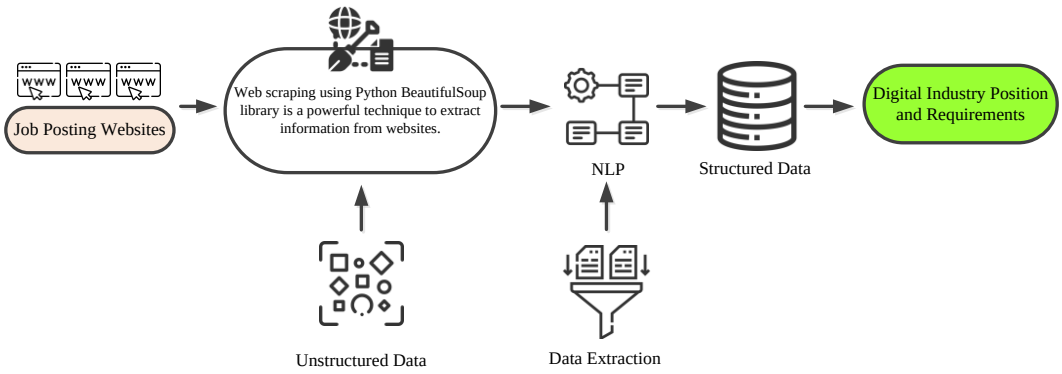

Figure 1. Proposed method framework

3.1. Web scraping techniques

Web scraping stands out as an effective technique for data collection from websites, leveraging powerful Python tools. The rationale for choosing this method lies in its versatility and efficiency, allowing researchers to quickly gather and process large volumes of data directly from the source in a programmable and scalable manner. This method not only reduces the manpower and financial resources needed for manual data collection but also enhances the accuracy and comprehensiveness of the data collected, making it invaluable for studies that require up-to-date and detailed information across various fields [17]. The primary steps involved in the data collection process are as follows:

3.1.1. Choose the job advertisement or posting website

The first step is to choose a website that specializes in job advertisements or postings. These websites typically host a plethora of relevant job postings, and identifying suitable platforms is generally straightforward. It is essential to ensure that these sites have a comprehensive and diverse range of listings that reflect the current job market trends in various industries. This approach not only streamlines the collection process but also guarantees that the data extracted is representative and extensive, facilitating a more accurate analysis of workforce requirements and employment opportunities across sectors. However, the criteria for selecting target websites must ensure that they feature job positions related to all five groups within the digital industry. This is crucial to ensure that the data gathered is comprehensive and effectively addresses the analysis of future workforce requirements. Selecting websites that meet this specific criterion helps to ensure that data collection is accurate and current, supporting the research and achieving clearly defined results.

3.1.2. Inspect the URL location

Proceeding to inspect the uniform resource locator (URL) location is crucial. Typically, the URLs on job listing webpages are in hypertext markup language (HTML). Ensuring that these websites do not employ scripts to impede data access is imperative. The inspection process involves utilizing web browser tools, such as the inspect method, to scrutinize the structure of the URL and confirm its accessibility for data extraction. To inspect with Google Chrome, right-click on the webpage and select the "Inspect" command to display the HTML. This step allows researchers to understand the underlying code and navigate potential barriers, such as dynamically loaded content via JavaScript, which can complicate data scraping. Figure 2 illustrates how the inspection window provides a clear view of the webpage's elements, highlighting the structure of data and its location for extraction, thereby facilitating a more targeted and efficient scraping strategy.

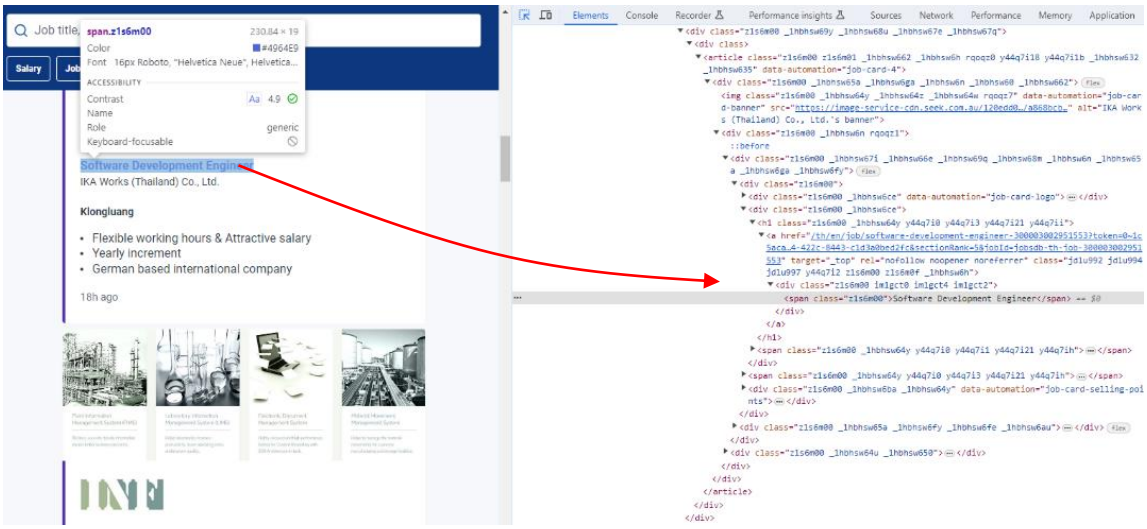

Figure 2. Inspecting URL location on a job list

3.1.3. Extracting all URLs

This step uses Python to systematically extract all URLs from the webpage containing job listings. This process uses the bs4 module from the BeautifulSoup library and stores the URL data in a Pandas data

frame [18]. We subsequently export the extracted data to Excel or CSV file formats to facilitate further analysis. Users can conveniently select specific groups of interest directly from the homepage thanks to the common categorization of job positions. Users can further streamline their search by selecting desired job position categories, enhancing the efficiency and precision of the data extraction process. Are visualized in Figure 3.

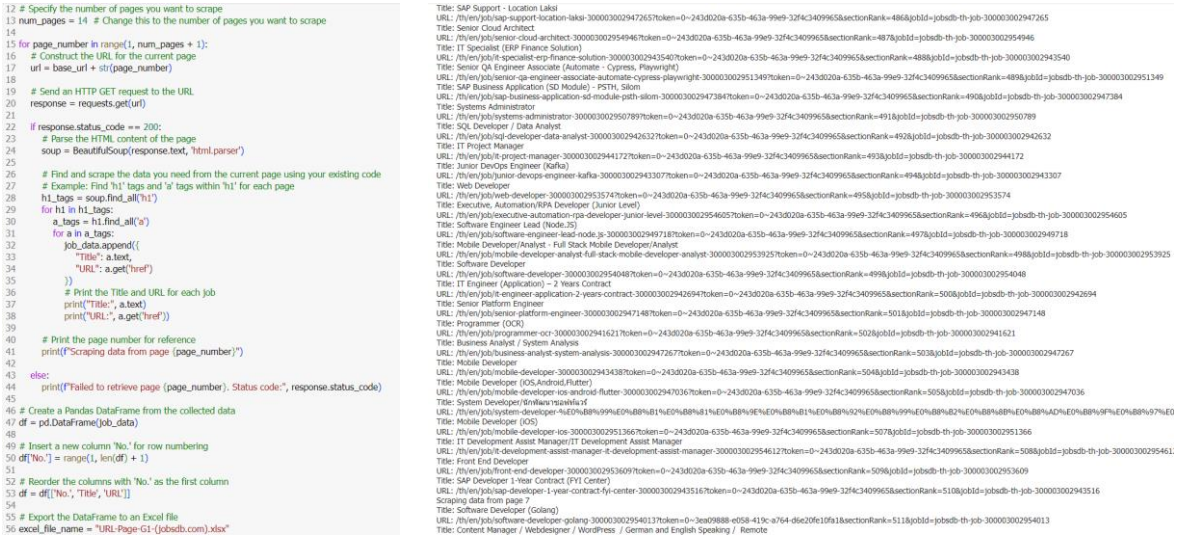

Figure 3. Extracting all URLs for job positions

3.1.4. Inspect the location of desired information

At this point, the focus is on the job title and description webpage, where HTML tags like headings (tag h1-h6), paragraphs (tag p), and lists (tag li) typically contain pertinent information. When inspecting the Google Chrome inspection page, carefully scroll through it to accurately locate the data [19]. It is essential to conduct a thorough examination, as certain tags might have specific IDs or classes, which adds an extra level of scrutiny for precise data extraction. This meticulous inspection ensures the precise location and retrieval of the desired information for comprehensive analysis. As depicted in Figure 4.

Figure 4. Inspecting the location of desired information

3.1.5. Web scraping using Python

For data aggregation, web scraping is a powerful tool that enables the collection and consolidation of information from diverse sources. Gathering vast amounts of data quickly and efficiently, and then

analyzing it to discern patterns, trends, or insights relevant to various industries or fields of study, is particularly effective with this method. At this stage, we execute web scraping using Python, utilizing the BeautifulSoup package library [20]. The extracted data is then exported in Excel file format. The methodology involves the efficient use of URL looping, enabling the simultaneous scraping of data from multiple web pages. This approach is particularly advantageous for handling large datasets, significantly optimizing time utilization [21], [22].

Given that the preceding step yields links to numerous job positions, the resulting text is often a mix of natural language, special characters, numbers, symbols, and other unstructured elements. Consequently, these texts will undergo NLP in the subsequent step to refine and structure the data for comprehensive analysis. This processing involves techniques such as tokenization, the removal of stopwords, and the extraction of key terms, which are essential for filtering out irrelevant information and highlighting significant details relevant to job market trends. As depicted in Figure 5, this step is crucial for converting raw data into a format that is both analyzable and actionable, ensuring that the insights derived are accurate and meaningful.

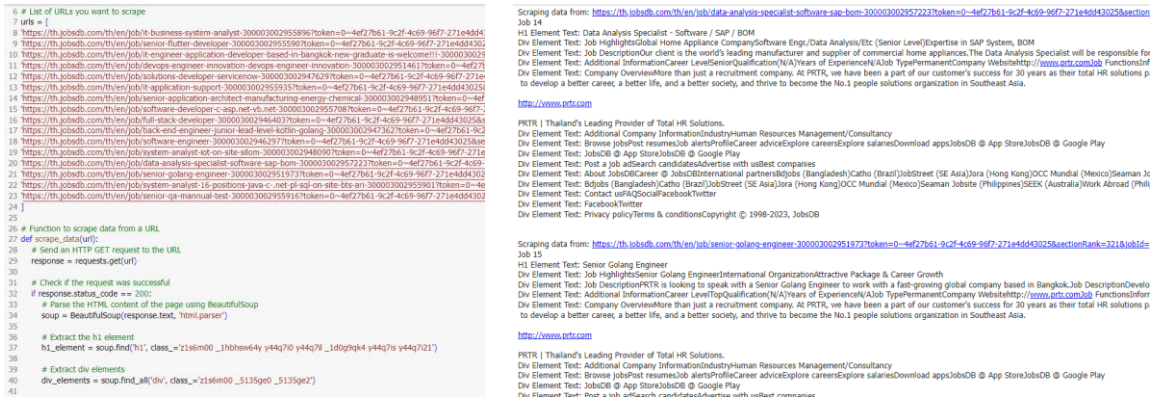

Figure 5. Web scraping operation

3.2. Natural language processing

In research studies, NLP is a pivotal process where textual information takes center stage. This technology bridges the gap between human communication and computer understanding, facilitating sophisticated interactions that allow for the extraction and analysis of key information from vast amounts of text. Computers can interpret, manipulate, and comprehend human language in this step, converting unstructured text into easily analyzed structured data. The process encompasses several key steps, as outlined as follows [23].

3.2.1. Data cleaning and normalization

This step entails fundamental data cleaning, such as removing redundant words or words that have a negligible impact on the analysis. The objective is to streamline and normalize the dataset, ensuring that it is devoid of unnecessary elements that might hinder the effectiveness of subsequent analyses. To achieve this, it is essential to identify and eliminate any duplicates, irrelevant terms, or placeholder text that could obscure meaningful insights. The key technique in this step involves removing redundant and irrelevant words for processing, using commands in Excel to find unwanted words and replace them with blank values. This process helps to maintain the integrity and quality of the data. Figure 6 illustrates this method.

3.2.2. Punctuation removal

This step focuses on cleaning and normalizing the text by removing punctuation marks. We systematically eliminate elements such as punctuation, misspellings, symbols, and other textual characteristics that could impact processing. This step contributes to refining the dataset for enhanced accuracy and ease of subsequent processing. Punctuation marks such as (!"#$%&'*+,-./:;<=>?@[]^_`{|}) are specifically targeted for removal. In this step, the technique involves importing the regular expression and natural language toolkit (NLTK) libraries in Python [24] to manage unwanted characters. Figure 7 illustrates this.

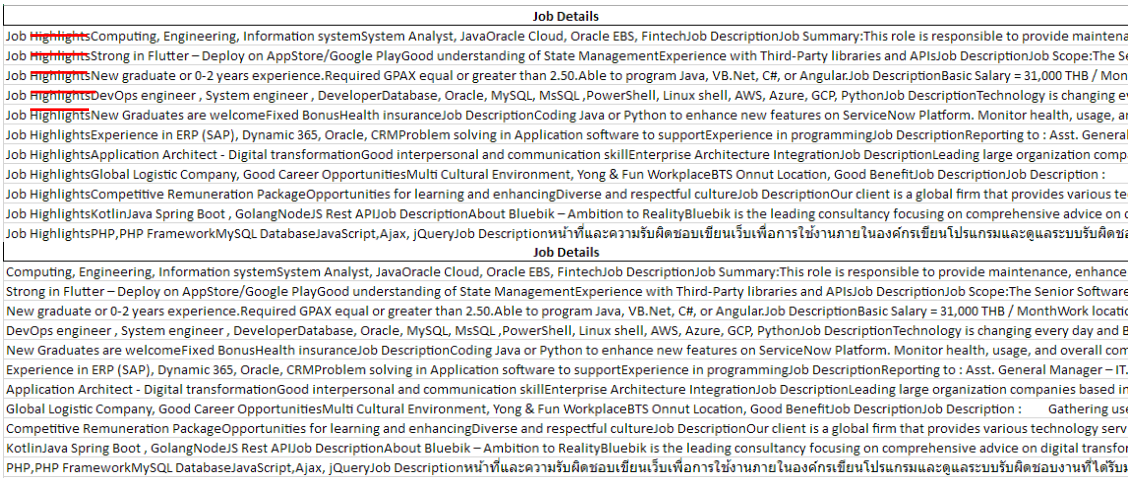

Figure 6. Data cleaning and data normalization

Figure 7. Punctuation removal operation

3.2.3. Tokenization

In this step, words undergo tokenization, a process that involves word segmentation and the removal of stop words. Tokenization breaks down the text into individual words, systematically eliminating stop words that can introduce noise in machine learning applications. We remove these superfluous terms, which we consider the cacophony of the text, to enhance the clarity and relevance of the analysis.

Notably, the removal of stop words contributes to obtaining meaningful nouns related to the job, streamlining the dataset by reducing the number of words. The tokenization process, which includes stop word removal, is executed using the Python NLTK and PyThaiNLP libraries, and custom additional stop words are incorporated to cater to the nuances of the Thai language [25]–[27]. Once we tokenize the text into words, we store the words in a Pandas data frame and export them to an Excel file for use in subsequent processes. As depicted in Figure 8.

3.2.4. Feature extraction

Feature extraction is a crucial step that involves studying job requirements based on the frequency of job positions. This analysis helps identify the most demanded skills and roles in the industry, which is particularly important for understanding employment trends and needs. We showcase the prominence of specific terms by creating a bag of words or a word cloud, which allows for a visual representation of key skills and requirements [28]. The research's results section will present the results of this feature extraction process, offering valuable insights into recurring job positions and their significance in the digital industry. This data is essential for organizations aiming to align their hiring strategies with market demands and for job seekers looking to enhance their qualifications in targeted areas.

Figure 8. Tokenization, word segmentation, and stop word removal operation

4. RESULTS

This study organizes its results into two primary segments: the first part analyzes job positions, highlighting the frequency and types of roles that are currently in demand. This section aims to highlight emerging trends and potential growth areas within the industry. The second part delves into the examination of workforce qualifications and requirements, assessing the specific skills and educational backgrounds employers are seeking. This analysis is crucial for understanding the alignment between job market demands and available workforce capabilities. The subsequent sections provide detailed insights into each facet of the study, offering a comprehensive overview that can inform stakeholders, including policymakers, educators, and industry leaders, in their decision-making processes.

4.1. Results of job positions

The study revealed numerous job positions with distinct requirements. To provide a comprehensive overview, we present the top 10 most common job positions within the digital industry, as outlined in Table 1. In addition to highlighting the most prevalent job positions in the digital industry sector, Table 1 serves as a crucial tool for understanding the core competencies and skills most valued in these roles. By analyzing these positions, we can discern the dynamic nature of the digital job market and the specific expertise that employers prioritize. This insight is vital for both job seekers aiming to enhance their careers and organizations striving to recruit the best talent in a competitive environment.

Examining job positions within the digital industry has revealed prevalent roles across several sectors. These include Electrical Engineers in hardware and smart devices, who design and develop new electronic equipment; Programmers in software and software services, who write and test code for various applications; IT support specialists in digital services, who maintain and troubleshoot information systems; graphic designers in digital content, who create engaging visual content; and network engineers in telecommunications, who construct and manage vital communication networks. These positions are in high demand across the job market, indicating a strong and ongoing need for a skilled workforce in these areas. Individuals with expertise in these domains are well-equipped to find stable and rewarding employment in the rapidly evolving digital industry [29].

Table 1. The most common job positions

No.	Hardware and smart device	Software and software service	Digital service	Digital content	Telecommunication
1	Electrical engineer	Programmer	IT support	Graphic design	Network engineer
2	Sales engineer	System analyst	Digital marketing	Creative	Telecommunications technician
3	Production engineer	Web developer	Application support	UX/UI design	Technical support
4	Network engineer	Full stack developer	Business analyst	Content creator	Production engineer
5	System engineer	Mobile app developer	Project manager	Video editor	Project engineer
6	Project engineer	System engineer	Data analyst	Tiktok content creator	Engineer specialist WAN
7	Service engineer	Software tester	Data engineer	Photographer	Service engineer
8	Maintenance engineer	Software developer	IT project management	Web designer	Signaling deliver engineer
9	IT support	Backend developer	Enterprise resource planning (ERP)	Video content creator	System engineer
10	Network operation center	Software engineer	Cloud engineer	Digital content creator	Communications specialist

4.2. Results of workforce requirements

The study looked into workforce qualifications based on job market demands within each digital industry group. This comprehensive analysis encompassed a range of criteria, including education level, skills, specialization, experience, and technological proficiency. To provide a broad perspective, we present the top 10 requirements that are most in demand, as outlined in Table 2. Table 2 not only sheds light on the key qualifications sought after in the digital industry but also emphasizes the evolving nature of these demands as the sector adapts to new technologies and market trends. The findings provide valuable insights for prospective workers seeking to align their career paths with market needs, as well as educational institutions seeking to update their curricula to better prepare students for future job opportunities.

Table 2. The most common qualifications and requirements

No.	Hardware and smart device	Software and software service	Digital service	Digital content	Telecommunication
1	Bachelor degree	Bachelor degree	Bachelor degree	Bachelor degree	Bachelor degree
2	Experience	Web development	Project manager	Graphic design	Telecommunication technical
3	Engineering	Mobile application development	Software service	Social media content	Maintenance
4	Customers	Software development	Customers service	Creative content	Vocational certificate
5	English skills	Technical support	English skills	Digital media	Electrical engineering
6	Electrical	System design	Data engineer	Advertising and Marketing	Customers service
7	Network	Advertising and Marketing	Customers service	Adobe Photoshop	English skills
8	Management	SQL database	Communication	Communication	Equipment skill
9	Technical	Project manager	Knowledge	English skills	Sales experience
10	Support	Software tester	Support	Adobe Premiere Pro	Products

Across all five industrial groups, the study on qualifications and workforce requirements reveals that educational qualifications are the primary prerequisites for workers, with a strong preference for individuals who have completed a bachelor's degree. Subsequently, deep knowledge, specialized abilities, and job-specific skills or functional competencies are of significant importance in securing employment in these sectors. Moreover, employers frequently seek out various other requirements, including the ability to work outside, availability for overtime, and proficiency in English. These findings underscore the holistic nature of the qualifications demanded in the digital industry, highlighting not only the technical skills but also the adaptability and soft skills that are critical for success. This comprehensive understanding helps guide job seekers in their career development efforts and assists educational institutions in tailoring their programs to meet the evolving demands of the workforce.

4.3. Results of natural language processing

The NLP phase's outcomes include word frequency analysis using the bag-of-words (BoW) technique and the creation of a word cloud for text feature extraction using Python. These results play a pivotal role in enhancing the understanding of textual data, offering a visual representation of the frequency and prominence of specific terms. We generate the word cloud using Python libraries like matplotlib and wordcloud to visually emphasize these frequencies [30], highlighting the most critical and frequently mentioned terms in an impactful and easily digestible format. The subsequent sections delve into the insights gained from these NLP analyses, providing a nuanced interpretation of the language used in the digital industry workforce requirements. This comprehensive approach aids stakeholders in aligning their strategies with the industry's actual needs by identifying key trends, common qualifications, and highly valued skills.

4.3.1. Job position analysis of Group 1

According to the word frequency analysis of job positions within digital industry Group 1, which focuses on hardware and smart devices, the top positions are as follows: i) electrical engineer (92 positions), ii) sales engineer (92 positions), and iii) production engineer (71 positions). Figure 9 depicts these positions and various others, offering a clear visual representation of the frequency distribution of job positions within this industry group. This visualization not only highlights the most in-demand roles but also illustrates the diversity of career opportunities available within the sector. Understanding these trends helps stakeholders, such as job seekers and educational institutions, align their goals and curricula with the industry's needs, ensuring they are preparing for the most prevalent and emerging roles.

4.3.2. Job requirements analysis of Group 1

The key requirements identified in the word frequency analysis of job requirements within digital industry Group 1, which specifically focuses on hardware and smart devices, are as follows: i) bachelor's degree (5,683 occurrences), ii) experience (3,766 occurrences), and iii) engineering (3,208 occurrences). Figure 10 visualizes these primary requirements and various others. Figure 10 provides a graphical representation of the frequency distribution of job requirements within this industry group, offering valuable insights into the qualifications sought by employers. This visualization helps to underscore the importance of formal education, practical experience, and specific engineering skills, guiding both job seekers in their career development and educational institutions in shaping their programs to meet industry demands.

Total number of words: 2022
Vocabulary size: 697

	Token	Count
0	Electrical Engineer	92
1	Sales Engineer	92
2	Production Engineer	71
3	Network Engineer	70
4	System Engineer	68
5	Project Engineer	60
6	Service Engineer	57
7	Maintenance Engineer	46
8	IT Support	40
9	Engineer	33
10	Network Operation Center	33

Figure 9. BoW and word cloud for job positions in Group 1

Total number of words: 292269
Vocabulary size: 35009

	Token	Count
0	Bachelor degree	5683
1	Experience	3766
2	Engineering	3208
3	Customers	2348
4	English skills	2182
5	Electrical	1989
6	Network	1843
7	Management	1711
8	technical	1238
9	support	1236
10	project	1191

Figure 10. BoW and word cloud for job requirements in Group 1

4.3.3. Job position analysis of Group 2

In the word frequency analysis of job positions within digital industry Group 2, focusing on software and software services, the top positions identified are as follows: i) programmer (244 positions), ii) system analyst (140 positions), and iii) web developer (106 positions). Figure 11 depicts these positions and various others. This visual representation provides a clear and comprehensive view of the frequency distribution of job positions in this industry group, offering valuable insights into the prevalent roles. The analysis not only highlights the demand for technical expertise in programming, systems analysis, and web development but also aids in understanding the evolving needs of the software industry, thereby assisting job seekers, educators, and industry leaders in making informed decisions.

Total number of words: 3335
Vocabulary size: 985

	Token	Count
0	Programmer	244
1	system Analyst	140
2	Web Developer	106
3	Full Stack Developer	103
4	Mobile App Developer	102
5	System Engineer	93
6	Software Tester	90
7	Software Developer	87
8	Back End Developer	86
9	Software Engineer	85
10	Front End Developer	61

Figure 11. BoW and word cloud for job positions in Group 2

4.3.4. Job requirements analysis of Group 2

In the word frequency analysis of job requirements within digital industry Group 2, which focuses on software and software services, the primary requirements identified include: i) a bachelor's degree (12,146 occurrences), ii) web development (9,316 occurrences), and iii) mobile application development (5,574 occurrences). Figure 12 visualizes these key requirements and various others. Figure 12 provides a clear and detailed view of the most sought-after qualifications by employers in the software and software services sector, offering a graphical representation of the frequency distribution of job requirements within this industry group. This analysis not only highlights the importance of formal education but also underscores the demand for specialized skills in web and mobile development, helping to guide both job seekers and educational institutions in aligning their efforts with market needs.

Total number of words: 418681
Vocabulary size: 40863

	Token	Count
0	Bachelor degree	12146
1	Web development	9316
2	Mobile Application dev	5574
3	Software development	4975
4	Technical Support	4062
5	System design	3568
6	SQL Database	2980
7	Project Manager	2803
8	Software Tester	2561
9	English skills	2526
10	Programming	1895

Figure 12. BoW and word cloud for job requirements in Group 2

4.3.5. Job position analysis of Group 3

In the word frequency analysis of job positions within digital industry Group 3, which encompasses digital services, the top positions identified are as follows: i) IT support (354 positions), ii) digital marketing (330 positions), and iii) application support (144 positions). Figure 13 visualizes these positions and various others. This visual representation provides insights into the frequency distribution of job positions within this industry group, offering a comprehensive understanding of the prevalent roles in the digital services sector. The analysis highlights the critical demand for support roles and digital marketing expertise, reflecting the sector's focus on maintaining operational efficiency and expanding online presence, thereby aiding stakeholders in targeting their career and business strategies effectively.

Total number of words: 4032
Vocabulary size: 1559

	Token	Count
0	IT Support	354
1	Digital Marketing	330
2	Application Support	144
3	Business Analyst	125
4	Project Manager	109
5	Data Analyst	74
6	Data Engineer	65
7	IT Project Management	55
8	ERP	49
9	Cloud Engineer	48
10	Database Administrator	46

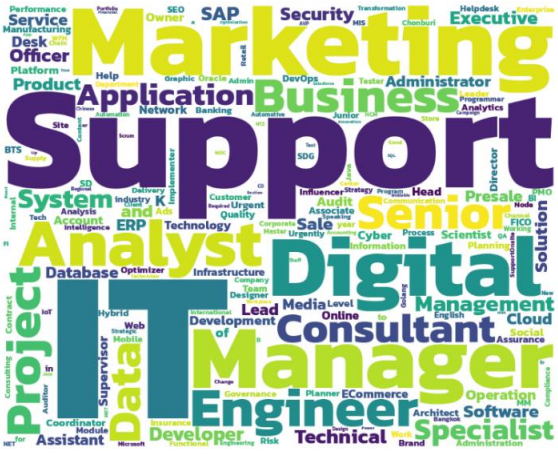

Figure 13. BoW and word cloud for job positions in Group 3

4.3.6. Job requirements analysis of Group 3

In the word frequency analysis of job requirements within digital industry Group 3, focusing on digital services, the main requirements identified are as follows: i) bachelor's degree (5,806 occurrences), ii) project manager (3,425 occurrences), and iii) software services (3,103 occurrences). These primary qualifications, along with other frequently mentioned criteria, are visualized in Figure 14. This illustration provides a graphical representation of the frequency distribution of job requirements within this sector, effectively highlighting the most sought-after qualifications by employers in the digital services industry. The analysis underscores the critical importance of formal education and specialized roles, such as project management and software services, guiding both educational institutions and job seekers in their strategic planning and career development efforts.

Total number of words: 225275
Vocabulary size: 24947

	Token	Count
0	Bachelor degree	5806
1	Project Manager	3425
2	Software Service	3103
3	Customers Service	2382
4	English skills	1661
5	Data Engineer	1100
6	Communication	1040
7	Knowledge	986
8	support	969
9	Database skill	966
10	Business	735

Figure 14. BoW and word cloud for job requirements in Group 3

4.3.7. Job position analysis of Group 4

In the word frequency analysis of job positions within the digital industry Group 4, which focuses on digital content, the top positions identified include i) graphic design (587 positions), ii) creative (148 positions), and iii) UX/UI design (125 positions). Figure 15 visualizes these positions and various others. Figure 15 provides a visual representation of the frequency distribution of job positions within this industry group, offering insights into the prevalent roles in the digital content sector. The analysis highlights the high demand for design and creative skills, reflecting the industry's emphasis on visual and UI aspects, which are crucial for engaging and captivating digital audiences. This information is invaluable for professionals looking to enter or advance in this dynamic field, as well as for educational institutions developing relevant curricula.

Total number of words: 2408
Vocabulary size: 809

	Token	Count
0	Graphic Design	587
1	Creative	148
2	UXUI Design	125
3	Content Creator	123
4	Video Editor	121
5	Tiktok Content Creator	46
6	Photographer	43
7	Web Designer	38
8	Video Content Creator	33
9	Digital Content Creator	32
10	Live Streaming	28

Figure 15. BoW and word cloud for job positions in Group 4

4.3.8. Job requirements analysis of Group 4

The primary requirements identified in the word frequency analysis of job requirements within the digital industry Group 4, focusing on digital content, are: i) a bachelor's degree (5,287 occurrences), ii) graphic design (3,719 occurrences), and iii) social media content (2,975 occurrences). Figure 16 visualizes these key qualifications and other frequently mentioned criteria. Figure 16 provides a graphical representation of the frequency distribution of job requirements within this industry group, offering insights into the qualifications most sought after by employers in the digital content sector. The analysis emphasizes the importance of formal education, specialized design skills, and expertise in social media content creation, highlighting these as crucial areas for professionals aiming to excel in this field. This information is invaluable for guiding educational institutions in curriculum development and assisting job seekers in aligning their skills with industry demands.

Total number of words: 141644
Vocabulary size: 20060

	Token	Count
0	Bachelor degree	5287
1	Graphic Design	3719
2	Social Media Content	2975
3	Creative Content	2486
4	Digital Media	2127
5	Advertising and Marketing	1898
6	Adobe Photoshop	1499
7	Communication	1461
8	English skills	978
9	Adobe Premiere Pro	789
10	Adobe Illustrator	609

Figure 16. BoW and word cloud for job requirements in Group 4

4.3.9. Job position analysis of Group 5

In the word frequency analysis of job positions within digital industry Group 5, focusing on telecommunications, the top positions identified include i) network engineer (37 positions), ii) telecommunications technician (32 positions), and iii) technical support (27 positions). Figure 17 visualizes these positions and various others. Figure 17 provides a visual representation of the frequency distribution of job positions within this industry group, offering insights into the prevalent roles in the telecommunications sector. The analysis highlights the critical need for technical expertise in networking,

telecommunications maintenance, and support services, reflecting the sector's reliance on skilled professionals to manage and maintain communication infrastructures. This information is essential for job seekers looking to enter or advance in this dynamic field, as well as for educational institutions preparing students for these critical roles.

Total number of words: 637
Vocabulary size: 338

	Token	Count
0	Network Engineer	37
1	Telecommunications Technician	32
2	Technical Support	27
3	Production Engineer	19
4	Project Engineer	19
5	Engineer Specialist WAN	15
6	Service Engineer	10
7	Signaling Deliver Engineer	10
8	System Engineer	10
9	Communications Specialist	9
10	Core Network Engineer	9

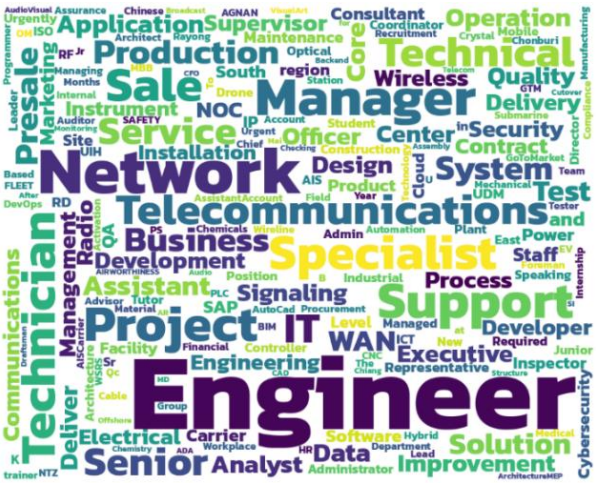

Figure 17. BoW and word cloud for job positions in Group 5

4.3.10. Job requirements analysis of Group 5

In the word frequency analysis of job requirements within digital industry Group 5, focusing on telecommunications, the primary requirements identified are as follows: i) bachelor's degree (167 occurrences), ii) telecommunication technical (88 occurrences), and iii) maintenance (81 occurrences). Figure 18 visualizes these key qualifications and various others. Figure 18 provides a graphical representation of the frequency distribution of job requirements within this industry group, offering insights into the qualifications sought by employers in the telecommunications sector. The analysis emphasizes the importance of formal education, technical expertise in telecommunications, and maintenance skills, reflecting the critical role of qualified professionals in ensuring the smooth operation of communication networks and services. This information serves as a valuable resource for job seekers and educational institutions seeking to align their efforts with the demands of the telecommunications industry.

Total number of words: 2994
Vocabulary size: 1364

	Token	Count
0	Bachelor degree	167
1	Telecommunication Technical	88
2	Maintenance	81
3	Vocational certificate	60
4	Electrical Engineering	55
5	Customers Service	39
6	English skills	38
7	Equipment skill	36
8	Sales Experience	32
9	Products	27
10	Safety	25

Figure 18. Bag-of-words and word cloud for job requirements in Group 5

5. DISCUSSION

The study of requirements for the digital industry workforce identified significant needs across five groups: i) In the hardware and smart devices group, the most required position is that of an electrical engineer; ii) Programmer is the most sought-after job title in the software and software services group; iii) In the digital service group, IT support is the most in-demand; iv) Within digital content, the role of graphic designer is most needed; and v) In the telecommunications group, the role of network engineer is highly sought after. Additionally, all industry groups require the workforce to have a bachelor's degree as a foundational skill.

However, when compared with previous studies, it appears that no research has yet clearly presented both the job positions and skill requirements demanded by the labor market. As a result, the findings of this study are beneficial for the workforce seeking employment in the digital industry, as they provide insights into the necessary skills for preparation and competition for the desired job positions. This new knowledge empowers the workforce to strategically position themselves within the competitive landscape of the digital industry, enhancing their employability and readiness to meet future challenges.

6. CONCLUSION

In conclusion, this research represents a systematic inquiry into the employment prerequisites of the digital industry, using web scraping and NLP techniques to provide valuable insights into job roles and requisite qualifications. The study's findings significantly contribute to the research topic by offering a detailed understanding of industry expectations, which is essential for both individuals aiming to enter the digital field and educational institutions striving to align their curricula with market needs. The positive impact of this research lies in its ability to bridge the gap between academia and the professional world, ensuring that graduates are well-equipped to meet the challenges and leverage the opportunities of the digital industry.

REFERENCES

[1] Ardi *et al.*, "The secret to enhancing innovativeness in digital industry," *Int. J. Innov. Creat. Chang.*, vol. 12, no. 12, pp. 225–243, 2020.

[2] N. Yung, "10 s-curve star industries it is expected that Thailand will take 5-10 years to understand technology and survive disruption." Marketing Oops!, 2018. Accessed: Oct. 20, 2023. [Online]. Available: https://www.marketingoops.com/news/biz-news/10-s-curve

[3] G. Gayatri, I. G. N. M. Jaya, and V. M. Rumata, "The Indonesian digital workforce gaps in 2021–2025," *Sustainability*, vol. 15, no. 1, p. 754, Dec. 2022, doi: 10.3390/su15010754.

[4] International Labour Organization, "Impact of digitalization on skills within and across sectors," in *Changing Demand for Skills in Digital Economies and Societies*, Geneva: ILO Publishing, 2021, pp. 62–88.

[5] S. Elhajjar, "Digital marketing jobs: what are recruiters looking for?," *J. Educ. Bus.*, vol. 97, no. 1, pp. 29–35, Jan. 2022, doi: 10.1080/08832323.2021.1884520.

[6] S. Tamara and S. Mike, "Attracting tomorrow's talent today: digital marketing strategies to recruit the workforce of the future," *J. Digit. Soc. Media Mark.*, vol. 6, no. 4, pp. 323–331, 2019.

[7] K. Jarmul and R. Lawson, "Introduction to web scraping," in *Python Web Scraping*, Birmingham: Packt Publishing, 2017, pp. 7–33.

[8] S. Lunn, J. Zhu, and M. Ross, "Utilizing web scraping and natural language processing to better inform pedagogical practice," in *2020 IEEE Frontiers in Education Conference (FIE)*, IEEE, Oct. 2020, pp. 1–9, doi: 10.1109/FIE44824.2020.9274270.

[9] J. Schedlbauer, G. Raptis, and B. Ludwig, "Medical informatics labor market analysis using web crawling, web scraping, and text mining," *Int. J. Med. Inform.*, vol. 150, p. 104453, Jun. 2021, doi: 10.1016/j.ijmedinf.2021.104453.

[10] V. Priya, N. Shri, M. V Janet, and M. Fathima, "Building a job portal using web scrapping," *Int. J. Progress. Res. Eng. Manag. Sci.*, vol. 3, no. 5, pp. 633–637, May 2023, doi: 10.58257/ijprems31247.

[11] L. MacKenzie, "Using web site scraping to match United States Virgin Island employers' job skill requirements with marketing curriculum and course development at the University of the Virgin Islands," *SSRN Electron. J.*, 2021, doi: 10.2139/ssrn.3787083.

[12] F. A. Riski, N. Selviandro, and M. Adrian, "Implementation of web scraping on job vacancy sites using regular expression method," in *2022 1st International Conference on Software Engineering and Information Technology (ICoSEIT)*, IEEE, Nov. 2022, pp. 204–209, doi: 10.1109/ICoSEIT55604.2022.10029964.

[13] K. Sinha, P. Sharma, H. Sharma, and K. Asawa, "Web scraping and job recommender system," in *2023 Second International Conference on Informatics (ICI)*, IEEE, Nov. 2023, pp. 1–6, doi: 10.1109/ICI60088.2023.10420941.

[14] C. Slamet, R. Andrian, D. S. Maylawati, Suhendar, W. Darmalaksana, and M. A. Ramdhani, "Web scraping and naïve Bayes classification for job search engine," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 288, p. 12038, Jan. 2018, doi: 10.1088/1757-899X/288/1/012038.

[15] S. Noordende and S. Scarpetta, "Digital skills: how businesses and policymakers can respond to future demand in the labour market." 2022. Accessed: Apr. 27, 2024. [Online]. Available: https://www.weforum.org/agenda/2022/11/digital-skills-labour-market-future

[16] Depa, "Thailand digital economy glance." World Economic Forum, 2017. Accessed: Sep. 05, 2023. [Online]. Available: https://www.depa.or.th/th/article-view/thailand-digital-economy-glance

[17] D. Glez-Peña, A. Lourenço, H. López-Fernández, M. Reboiro-Jato, and F. Fdez-Riverola, "Web scraping technologies in an API world," *Brief. Bioinform.*, vol. 15, no. 5, pp. 788–797, Sep. 2014, doi: 10.1093/bib/bbt026.

[18] D. Albuja, L. Guerra, D. Rivero, and S. Quishpe, "Optimizing marketing through web scraping," in *Smart Innovation, Systems and Technologies*, Springer, Singapore, 2023, pp. 739–755, doi: 10.1007/978-981-19-9099-1_51.

[19] A. Isasi-Andrieu, E. Garrote-Contreras, P. Iriondo-Bengoa, D. Aldama-Gant, and A. Galdran, "Deflectometry setup definition for automatic chrome surface inspection," in *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, IEEE, Sep. 2017, pp. 1–4, doi: 10.1109/ETFA.2017.8247756.
[20] Styawati, A. Nurkholis, F. A. Ans, S. Alim, L. Andraini, and R. A. Prasetyo, "Web scraping for summarization of freelance job website using vector space model," in *2023 IEEE 9th Information Technology International Seminar (ITIS)*, IEEE, Oct. 2023, pp. 1–5, doi: 10.1109/ITIS59651.2023.10420412.
[21] A. Abodayeh, R. Hejazi, W. Najjar, L. Shihadeh, and R. Latif, "Web scraping for data analytics: a beautifulsoup implementation," in *2023 Sixth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU)*, IEEE, Mar. 2023, pp. 65–69, doi: 10.1109/WiDS-PSU57071.2023.00025.
[22] S. Kumar and U. B. Roy, "A technique of data collection: web scraping with python," in *Statistical Modeling in Machine Learning*, Elsevier, 2023, pp. 23–36, doi: 10.1016/B978-0-323-91776-6.00011-7.
[23] C. P. Chai, "Comparison of text preprocessing methods," *Nat. Lang. Eng.*, vol. 29, no. 3, pp. 509–553, May 2023, doi: 10.1017/S1351324922000213.
[24] F. Millstein, *Natural language processing with python: natural language processing using nltk*. Frank Millstein, 2020.
[25] N. Hardeniya, "Text wrangling and cleansing," in *Natural Language Processing: Python and NLTK*, Birmingham: Packt Publishing, 2016, pp. 21–30.
[26] W. Phatthiyaphaibun *et al.*, "PyThaiNLP: thai natural language processing in python," *arXiv*, Dec. 2023, [Online]. Available: http://arxiv.org/abs/2312.04649
[27] W. Chansanam, K. Tuamsuk, K. Kwiecien, and S. G. Oh, "Korean-popular Facebook fan page analytics in Thailand," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 5, pp. 5291–5303, Oct. 2022, doi: 10.11591/ijece.v12i5.pp5291-5303.
[28] Y. Jin, "Development of word cloud generator software based on python," *Procedia Eng.*, vol. 174, pp. 788–792, 2017, doi: 10.1016/j.proeng.2017.01.223.
[29] R. Rikharom and W. Chansanam, "Essential competencies for digital workforce of provincial office in Thailand using Delphi technique," *J. Inf. Sci. Theory Pract.*, vol. 11, no. 4, pp. 51–81, 2023, doi: 10.1633/JISTaP.2023.11.4.4.
[30] P. Zhang and Y. Pan, "A comparative study of keywords and sentiments of abstracts by python programs," *Open J. Mod. Linguist.*, vol. 10, no. 06, pp. 722–739, 2020, doi: 10.4236/ojml.2020.106044.

BIOGRAPHIES OF AUTHORS

Rachawit Tipsena is a Ph.D. student in information science. Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand. Skilled in data mining, text mining, and machine learning. Graduated with a master's degree in information technology from Mahasarakham University, Thailand. He can be contacted at email: rachawit.t@kkumail.com.

Wirapong Chansanam is an experienced Associate Professor with a demonstrated history of working in information science. He earned his Ph.D. in Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand, in 2014. Currently, he serves as the Head of the Information Science Department and Chair of the Digital Humanities Research Group at Khon Kaen University. His research interests include information sciences, ontology, knowledge organization systems, and linked open data. He can be contacted via email: wirach@kku.ac.th.

Kulthida Tuamsuk has a Doctor of Arts in Library and Information Science from Simmons University, USA. She is a professor and senior researcher in the Department of Information Science, Faculty of Humanities and Social Sciences, and a director of the Smart Learning Innovation Research Centre at Khon Kaen University, Thailand. Her research interests are knowledge management, digital humanities, metadata, ontology, digital learning, and smart learning. She can be contacted via email: kultua@kku.ac.th.

Tossapon Boongoen has a Ph.D. in Computer Science from Cranfield University, UK (in 2003), and his 2-year PostDoc at Department of Computer Science, Faculty of Business and Physical Sciences, Aberystwyth University, UK (2007-08). Before his current appointment, he worked for Mae Fah Luang University (MFU, Thailand) between 2017-2022 as an assistant professor at the School of Information Technology and the director of MFU Research and Innovation Institute. His research areas include AI, machine learning, uncertainty and fuzzy systems, data science, and engineering. He has also served as associate editor of international journals like IEEE Access, PeerJ Computer Science, ICT Express, and International Journal of Artificial Intelligence. He can be contacted at email: tob45@aber.ac.uk.