Comparative Analysis of Fast Fashion and Slow Fashion Website Design Elements Using a Web Crawler

University of California, Irvine

Grace Manning

June 2023

Acknowledgement

I would like to thank Dr. Alberto Krone-Martins for his ongoing mentorship and support through the research process. I would also like to acknowledge the UC Irvine Campuswide Honors Collegium for their encouragement to take part in undergraduate research.

Abstract

Fast fashion and slow fashion retail websites are often visually distinct, but little research has been done to analyze statistical differences in their structure and content. As the conversation about the social and environmental impacts of fast fashion continues to grow, the need for quantitative comparisons between fast and slow fashion also rises. This research aims to collect quantitative data from the retail websites of five fast fashion brands and five slow fashion brands in an effort to identify and analyze their differences. To collect data, this project utilized a web crawler; a crawler is an automated program that can visit a website, collect its content, and return this information for analysis. The collected content includes hypertext markup language (HTML), which is the basic programming language for creating web pages. The crawler collected data on the websites of five fast fashion brands: Forever21, SHEIN, Fashion Nova, H&M, and PrettyLittleThing, and five slow fashion brands: Big Bud Press, CHNGE, Fashion Brand Company, Tunnel Vision, and iGirl. Overall, fast fashion brands had a larger number of pages (over 26,000) and categories (over 300), yet remained similar to slow fashion brands in other factors. This data scrapes the surface of what can be used to help consumers identify a fast or slow fashion brand solely based on website data. Additionally, it can be hard to understand the negative impacts that fast fashion has on our lives. These numbers help us quantify the overproduction and waste that comes from following trend cycles and buying into fast fashion. The use of a crawler to collect data can be expanded in many more ways to continue learning from retail websites.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The fashion industry is dominated by fast fashion companies, who utilize a rapid design and production cycle that follows current fashion trends and allows them to offer clothing and other goods at a low price. One of the leading sources on fashion brand ratings, Good On You, defines fast fashion as "cheap, trendy clothing that samples ideas from the catwalk or celebrity culture and turns them into garments in high street stores at breakneck speed to meet consumer demand" [18]. For example, the brand Zara can complete the cycle from design conception to a ready-for-sale retail product in just 15 days [10]. This quick process often utilizes cheap or unethical labor and unsustainably sourced materials, and can lead to overproduction and excessive landfill contributions. On the opposite end of the spectrum, slow fashion companies attempt to combat these harmful business practices. The slow fashion movement is an approach which considers the processes and resources required to produce clothing. It focuses on sustainability, quality garments, and ethical treatment of people, animals, and the planet [17].

In recent years, e-commerce has grown to be incredibly important for fashion retailers. The online retail industry is expected to reach a valuation of $672.71 billion by 2023 [2], including sales of clothing, footwear, and other accessories. This growth has been partially attributed to an increase in access to the Internet and a growing use of smartphones [7]. Therefore, a powerful website is important for companies who want to capitalize on this growing market.

As websites are becoming increasingly important in the success of fashion retailers, there is a growing interest in compelling website design and composition. Website design varies as brands aim to achieve different goals, such as influencing users to spend money or focusing on a positive user experience. Fast fashion websites are often packed with lots of visual elements which may excite their users and point them towards new collections or sale items. An example of this can be seen in SHEIN's website, which is modeled in Figure 1.1. Slow fashion websites have some similarities, such as displaying current promotions in eye-catching graphics. However, they are often less visually overwhelming or feature fewer elements on the page at one time. The website of a slow fashion brand, iGirl, is modeled in Figure 1.2 below. While they are visually contrasting, is there a significant quantifiable difference in how they are designed and built?

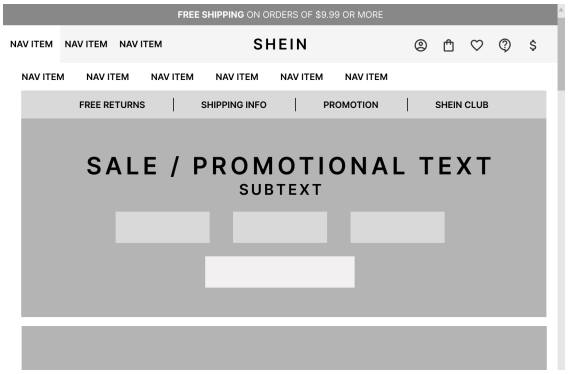

Figure 1.1: Layout style of Shein's website landing page on October 30, 2022.

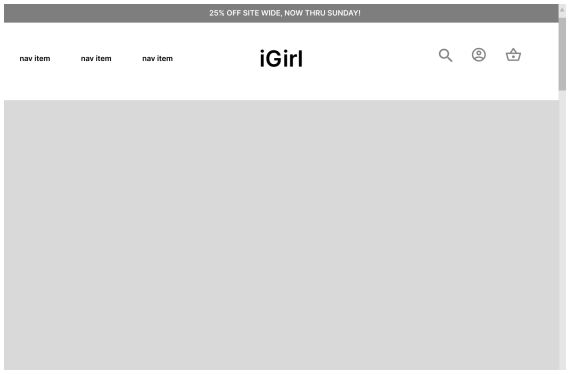

Figure 1.2: Layout style of iGirl's website landing page on October 30, 2022.

This research aims to answer the guiding question: are there explicit differences in the website design and components of a fast fashion website compared to a slow fashion website? A web crawler will be utilized to scrape data from fast fashion and slow fashion websites. For example, number of product collections, mention of sustainability certifications, and number of total pages associated with a website domain. This data will then be analyzed to determine if there are significant patterns and/or explicit differences between the two types of websites.

The purpose of this research is to understand the differences, if they exist, between fast

fashion and slow fashion website design and composition. This may be useful for further research, such as observing how website design and psychology are intertwined [19], or informing consumers of the type of brand they are shopping from. Website composition and design elements may be used to encourage people to spend money, while other elements may contribute towards a brand's image to appear more sustainable or ethical. Is it possible to determine if a brand is fast or slow fashion simply using data points captured from their websites? Before answering those questions, we must first determine if there is any quantifiable difference between fast fashion and slow fashion websites.

Chapter 2

Background

Previous research has visually analyzed retail website design, focusing primarily on fast fashion brands and the impact of interface design on consumers. In 2013, a group of researchers studied the design of websites for H&M, UNIQLO, and ZARA, finding that their websites all featured links to social networking services, displayed a variety of promotional videos, and released new items every week [13]. When looking at multiple website design parameters, other research has found that multiple parameters must exist to create a positive user experience; there is a strong relationship between security and prestige, as well as usability and prestige [6]. However, design features such as layout, graphics, and text were found to be less important than features like accessibility, speed, and navigation [6].

The relationship between consumers and vendors through websites has also been heavily studied; it was found in 2010 that consumers generally trust e-commerce sites, although that can be negatively impacted by lack of functionality, normality, and quality [15]. Additionally, brands are more likely to be recommended by word-of-mouth if their websites provide clear and accurate product information, while aesthetics showed a positive effect on consumer trust [19]. As people of all age groups are growing to value a circular fashion economy [9], it is increasingly important to consider slow fashion brands in future research of website design. Thus, my research builds on these works by analyzing the source code and underlying composition of retail websites while making a comparison between fast and slow fashion brands.

One way to collect large-scale information from website pages is through web crawlers, which are programmed bots that systematically browse the Internet. Crawlers start their search with a set of seed URLs, which is the starting set of URLs that the crawler initially visits. From there, the crawler can be customized to collect other links on those websites. Crawlers can return the raw HTML source code of a website, and afterward, various tools can be constructed to parse the retrieved code into meaningful pieces of information.

Web crawling etiquette is also essential to consider when using crawlers since websites are created for humans and not for other machines or robots. Many websites have a robots.txt file (located by default at addresses such as `www.domain-name.com/robots.txt`), which describes rules that crawlers should follow; ignoring it runs the risk of a crawler, and the physical machine from where the crawler was launched, being banned from the site entirely. For example, the robots.txt file commonly blocks pages such as the login, sign-up, shopping cart, and checkout pages on a retail website. This disallows a robot from creating an account, adding items to a cart, and purchasing those items. The crawler used for this research is aware and respectful of the rules laid out in each website's robots.txt file. Despite following the rules, some websites have automatic bot detection enabled and will quickly block robots as soon as the activity is found. To avoid raising any red flags with bot detection systems, crawlers can be built to utilize a randomly-generated delay that mimics human behavior.

Chapter 3

Methods

3.1 Exploration of Tools, Languages, and Supporting Technologies

Before crawling any websites, initial research was done to explore the different types of crawler technologies that currently exist. Factors considered for a crawler were the system(s) it can run on, the language(s) it works with, its mutability, and the amount of documentation provided to work with it. Crawlers like Scrapy and Selenium run using Python, whereas Heritrix uses Java and primarily runs on Linux systems. Ultimately, Headless Chrome Crawler was selected because it seemed most appropriate for this web-based research; it runs on Chrome and works with JavaScript and jQuery commands, which are commonly used in web development. It is described as being a "distributed crawler powered by Headless Chrome," meaning it can run Chrome without actually opening up a Chrome window for a human to view. However, the documentation for Headless Chrome Crawler is limited, and it is not used on a wide scale, so there are fewer online examples for its use compared to other technologies. This generated a larger learning curve and posed an issue for customizing it later on.

Determining how to store the data for this research was a crucial next step. Popular databases like Oracle XE and PostgreSQL were considered because they use the popular

query language SQL, along with being free and easy to use. However, the Redis database was eventually selected because it uses a key-value pair system to store data, which follows the simplicity of the data used in this research. The development of this project was done on a Windows machine, and Redis is not supported on Windows. To circumvent this issue, Windows Subsystem for Linux (WSL) was installed and run via the Windows command-line, at which point the Redis database could be launched for local development and testing.

The initial code for this research was built using the Headless Chrome Crawler library with a JavaScript codebase. This project utilized Node.js, a JavaScript runtime environment, and npm, or "node package manager," which allows JavaScript developers to install or develop packages for their projects. Headless Chrome Crawler is an example of an npm package. The codebase version control of this project was managed via a GitHub repository.

After initial setup and testing of the crawler and database, some issues were found. First, it was difficult to find resources online for errors with Headless Chrome Crawler. Many resources referred to Puppeteer, another crawler technology which Headless Chrome Crawler is built upon. Due to the more in-depth documentation and online resources associated with Puppeteer, the decision was made to switch to this crawler technology. The second issue found in the initial stage of this project was that JavaScript did not provide enough support needed to interact with the Redis database. Due to the nature of Redis and its use of data types, it was important that data types were well represented in the code. Thus, a switch was made to use TypeScript, which supports syntax for strongly-typed programming in JavaScript. The project built on Headless Chrome Crawler and JavaScript was named version 1, and that which utilized Puppeteer and TypeScript was named version 2. Moving forward, any mention of the crawler or the codebase refers to version 2.

Lastly, this project utilized a virtual machine from Amazon Web Services (AWS) called Elastic Compute Cloud (EC2). Running a crawler on large websites can take many hours or days, so it is essential to run it on a remote virtual machine rather than a local device. Thus, the crawler and Redis database were both deployed on the EC2 for data collection.

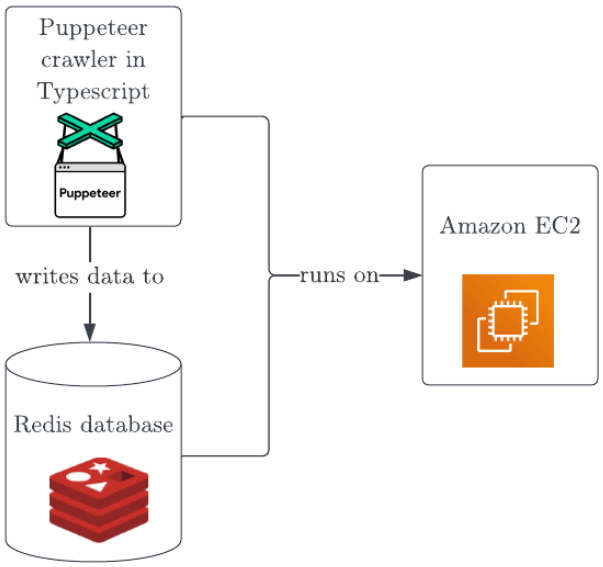

Figure 3.1: Crawler technology

For more information about the crawlers, databases, languages, and other technologies discussed in this section, as well as the codebase used in this project, see Appendix A: Technology Reference.

3.2 Sample Selection

Before writing the crawler, ten brands were selected for their websites to be examined. The sample was broken into five fast fashion brands and five slow fashion brands. To select these brands, it was important to start by understanding the biggest players in the fast fashion game. The website Good On You lists brands like CIDER, Fashion Nova, and SHEIN as fast fashion brands they would avoid shopping from due to ethical and environmental concerns [12]. Other sources, such as an article in Bloomberg magazine, named Shein Group, H&M, Zara, and Boohoo Group as some of the top fast fashion offenders for pollution and other environmental or ethical issues [5]. Ultimately, a list was created to accurately represent some of the most frequently mentioned fast fashion brands: Forever21, SHEIN, Fashion Nova, H&M, and PrettyLittleThing.

Selecting the fast fashion brands was a relatively quick process because they encompass the largest and most dominant brands in the fashion retail industry. Determining which slow fashion brands to use was a more challenging process because the slow fashion term refers to much smaller businesses that are often less known and harder to find. Crowd-sourcing became the best option to pick the five slow fashion brands: checking online resources, asking friends, and drawing from personal experience to determine which brands were most often mentioned and widely regarded as sustainable companies. Each company was explored further to confirm their status as a slow fashion brand, such as investigating their published transparency regarding the sources of their fabric, materials, and labor. Similar to the selection of fast fashion brands, a list was compiled of the suggested slow fashion brands, making sure to include the companies that were mentioned more than once. The brands selected were: Big Bud Press, Sézane, Fashion Brand Company, Tunnel Vision, and iGirl.

The websites needed to meet these criteria: they must allow crawlers to access their website, and the crawler must be able to gather valuable content. Many websites use CAPTCHA, or "Completely Automated Public Turing test to tell Computers and Humans Apart." These are tests that can be used to detect and block robot activity, including crawlers, on websites. After testing a basic crawler on all ten retail websites, there appeared to be an issue with one: Sézane used a CAPTCHA system that blocked all bot access. Since there is no legal workaround to bypass CAPTCHA technology (except to purchase credits from a third party that uses humans or machine learning to pass the tests), Sézane needed to be replaced. The next company on the list of slow fashion brands was selected, CHNGE, which passed the basic crawler tests. Thus, the revised list of slow fashion brands was Big Bud Press, CHNGE, Fashion Brand Company, Tunnel Vision, and iGirl.

It is important to note that this research focused on United States domains. When accessed by a user, retail websites determine the region from which the user is browsing. This region is used to determine the language and currency displayed on the website, which is often done using a different domain. For example, accessing SHEIN from the United States

yields `www.us.shein.com`, but users from the United Kingdom land on `www.shein.co.uk`. To maintain a manageable sample size, this research was limited to United States domains only. For a full list of brands and their corresponding websites crawled, see Table 3.1.

Brand	Website	Category
Forever21	`https://www.forever21.com`	fast fashion
SHEIN	`https://us.shein.com`	fast fashion
Fashion Nova	`https://www.fashionnova.com`	fast fashion
H&M	`https://www2.hm.com/en_us/index.html`	fast fashion
PrettyLittleThing	`https://www.prettylittlething.us`	fast fashion
Big Bud Press	`https://bigbudpress.com`	slow fashion
CHNGE	`https://chnge.com`	slow fashion
Fashion Brand Company	`https://www.fashionbrandcompany.com`	slow fashion
Tunnel Vision	`https://shoptunnelvision.com`	slow fashion
iGirl	`https://igirlworld.com`	slow fashion

Table 3.1: Brands and their websites used for crawling.

3.3 Data Collection

When determining which data points to collect from the retail websites, the goal of utilizing a crawler for data collection remained at the forefront. Thus, the data points needed to be within the page metadata, page content, page URL, or other available information. The initial brainstorming of possible data to collect from the websites included the number of links in the menu / navigation bar, number of size and color options, product prices, and amount of products displayed per page. Crawlers can collect an overwhelming amount of data, so this project primarily focused on website content that may verify sustainability practices or point to greenwashing[1]. Due to the nature of this research, it was decided that the crawler would count the mention of certifications that verify the labor, textile, and manufacturing standards of a business, as well as keywords like "sustainability" or "recycled."

[1]Greenwashing is the act of deceitfully advertising an organization's environmental initiatives or investing more resources in portraying the organization as eco-friendly than in genuinely implementing environmentally sustainable practices [4].

Exploring the capabilities of the Puppeteer crawler revealed new data points that were not initially considered. For example, it collected every cookie on a website, including metadata pertaining to the cookies. Each cookie consists of a name, value, domain (the website using the cookie), path (the website path using the cookie), expiration time, size, and other data. Additionally, the number of pages were counted for each website to understand the vast quantity of products listed, along with other content, such as blogs, hosted on a domain. The time to crawl was also logged for each website, although this was impacted by many factors, including the size of the website, whether the crawler needed to bypass Cloudflare or another bot detection service, and the strength of the Internet connection the crawler used to access a website.

The finalized set of data points collected by the crawler were: keywords, certifications, number of pages, categories, and cookies. See Table 3.2 for a more detailed breakdown of these data points.

Data Point	Explanation
keywords	Website content was searched for the following keywords: sustainability, sustainable, recyclable, recycled, reusable, environment, environmentally friendly, social responsibility, ethical, fairtrade.
certifications	Website content was search for the following certifications: B Corp / B-Corp, OEKO TEX / OEKO-TEX, Supplier Ethical Data Exchange (SEDEX), Sociéte Générale de Surveillance (SGS), Worldwide Responsible Accredited Production, North American Free Trade Agreement (NAFTA), Global Organic Textile Standard (GOTS), Leadership in Energy and Environmental Design (LEED).
number of pages	The number of pages refers to the pages within a website domain. For example, if a URL is www.example.com/about, the domain is example.com and the page is /about.
categories	Categories, also called collections, are often found within the menu or navigation bar on a website.
cookies	Cookies are used by sites to remember information about a user.

Table 3.2: Data points collected and their corresponding explanations.

3.4 Building the Crawler

The project began with a basic Puppeteer crawler, written using TypeScript. The crawler used a seed or set of seeds, which can be thought of as starter URLs, to begin crawling. As it traversed each website, the crawler collected links and added these to a queue. It traversed one at a time until the queue was empty. Throughout this process, the crawler also kept track of the links it had seen. If it had already seen a link, it was not re-added to the queue. Figure 3.2 provides a visual representation of this process. To improve the efficiency of this process, crawlers may implement a similarity detection system. For this research, a SHA-1 hash was utilized for exact similarity detection. SHA-1 (Secure Hash Algorithm 1) is a hash function that takes an input and creates a unique value, often represented as 40 hexadecimal digits [3]. If the content of two website pages generate the same SHA-1 value, then they are exactly the same. For this research, the crawler maintained a list of SHA-1 hash values generated for each visited website; if a website generated a hash value that was already in the list, the website was skipped. This prevented the crawler from scraping the same content more than once.

To circumvent the bot detection systems employed by modern websites, a delay was implemented in the crawler. After initially accessing a website, the crawler must wait for a randomly-generated time between 1000 and 2000 milliseconds before scraping the page for data. Using a different delay each time helps the crawler emulate human behavior. Additionally, changing the user agent for the crawler was extremely effective in avoiding bot detection. In HTTP requests, the user agent provides servers and peers with information to identify the application, operating system, vendor, and/or version of the requesting user [14]. In Puppeteer, the *setUserAgent* method takes in a string parameter and changes the user agent accordingly. Lastly, some fast fashion websites used Cloudflare bot detection which blocked the crawler from accessing their content. To fix this, the puppeteer-extra package was used to modify the Puppeteer crawler with the stealth plugin. The stealth plugin helps

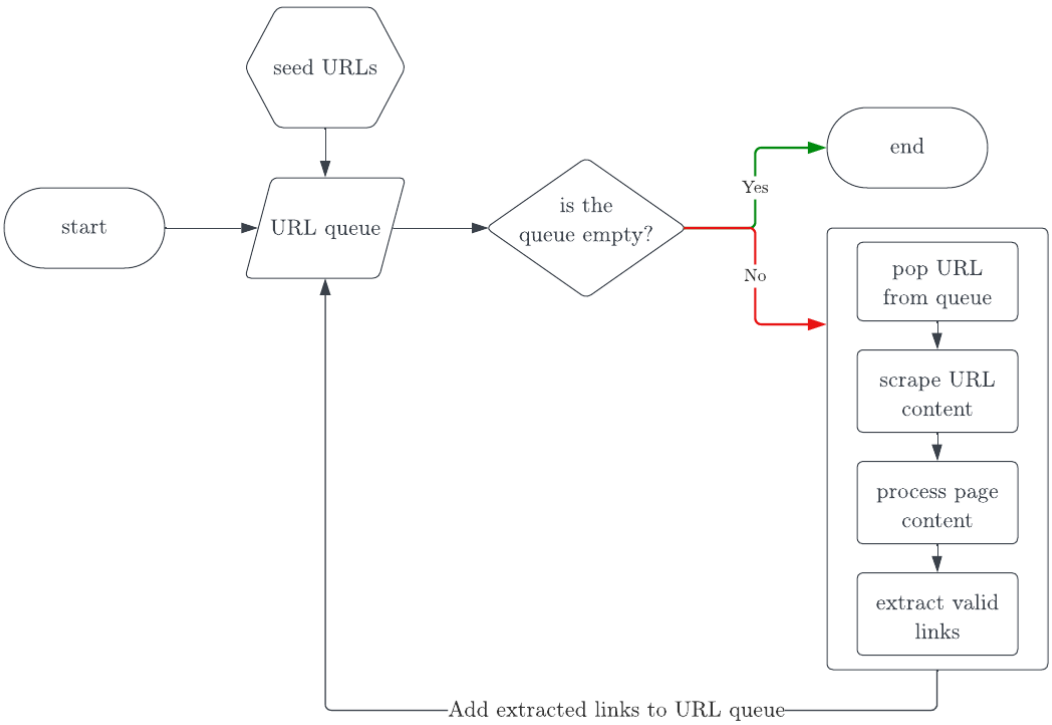

Figure 3.2: Crawler process diagram.

prevent detection for Puppeteer-based crawlers. After implementing it, there were no more issues with Cloudflare or any other detection service.

After handling bot detection, the crawler was further developed using the Puppeteer API to collect data points from the scraped content. Each page was checked for a 404 status code to ensure it loaded properly. The *evaluate* method of the Puppeteer API allows for interaction with the content of the page as an HTML Document. This method returns a Document object, so methods such as *getElementsByTagName* were used to extract all links from the page. The crawler also utilized Puppeteer API functions, including the *cookies* method, which returns the cookies a website uses, or the *content* method, which returns the content on a page that can be scraped for keywords and certifications. To count the categories on a website, URL processing was used to recognize the patterns that each retail brand uses to list categories on their website. For example, on Shopify websites, the categories can be found via URLs of the format "`www.example.com/collections/category-name`."

Due to the size of the fast fashion companies, it was beyond the scope of this project to let the crawlers traverse the entirety of their site. Thus, the crawler was force-stopped after 15 hours of crawling on one website. Some data collection was incomplete for this reason, such as the count of total pages on a website. To collect an accurate number of pages, a simple crawler was built to traverse the links listed on the sitemap page for a brand's website. A sitemap is a list of pages that exist on a website. This can often be found at `www.example.com/sitemap.xml`, but many large websites use a `sitemap_index.xml` file instead. The correct sitemap for a website can be found on the website's `robots.txt` page. If a sitemap index is listed, that index will contain links to sitemaps that each contain a subset of links to the website's pages. Figure 3.3 shows the sitemap index of fast fashion brand Forever21, while Figure 3.4 displays that of slow fashion brand Big Bud Press.

Figure 3.3: Forever21.com's sitemap index, which lists where to find other sitemaps. This page can be found at `forever21.com/sitemap_index.xml`.

Sitemaps are written in XML, a markup language standard for representing information that is commonly used for exchanging data over the Internet [16]. XML format uses a tag structure similar to HTML, in which data is contained within an opening and closing tag, denoted by `< >` and `</ >`. Some tags used in sitemap files are `<sitemap>` to denote other sitemap files, `<loc>` to list the location (link) of a page, and `<lastmod>` to inform of the last time the page was modified. The sitemap files are automatically updated when pages are added or deleted for websites hosted on Shopify, meaning they provided an extremely

accurate page count. Other website infrastructures may have automatically- or manually-updated sitemaps. The accuracy of a sitemap can also be inferred from its last modified date, which shows how often the sitemaps are maintained. To accurately count the number of pages within a website, the sitemap crawler visited the sitemap index files, collected links to the sitemaps, then traversed each sitemap while counting the number of links listed using the `<loc>` data. By counting each link listed on the sitemaps, this provided an accurate count of the number of pages on a website.

Figure 3.4: BigBudPress.com's sitemap index, which lists where to find other sitemaps. These sitemaps are categorized by links to products, pages, collections, and blogs. This page can be found at `bigbudpress.com/sitemap.xml`.

Chapter 4

Results

The crawler results show that both fast and slow fashion brands put a similar amount of sustainability-related keywords on their websites. The highest count of keywords was a total of 6 from Forever21, a fast fashion brand. Fashion Nova and PrettyLittleThing (fast fashion) and Tunnel Vision (slow fashion) had 0 keywords. The average number of keywords found on fast fashion sites was 4 keywords, and the average for slow fashion sites was 2.4 keywords. See Table 4.1 for a breakdown of the keywords found on each website.

Brand	Keywords	Count
Forever21	environment, environmentally friendly, ethical, recyclable, recycled, social responsibility	6
SHEIN	environment, social responsibility, sustainability, sustainable	4
Fashion Nova	-	0
H&M	environment, ethical, recycle, sustainability, sustainable	5
PrettyLittleThing	-	0
Big Bud Press	ethical, recycled, sustainable	3
CHNGE	fairtrade, environment, recycled, recyclable, sustainability	5
Fashion Brand Company	environmentally friendly, ethical, recyclable	3
Tunnel Vision	-	0
iGirl	ethical	1

Table 4.1: Keywords found on each website.

The crawler found that the fast fashion and slow fashion websites examined have a similar number of certifications listed on their websites. The average number of certifications on the

fast fashion websites was 1.4, while the average for slow fashion was 2.2. One slow fashion website, Fashion Brand Company, had 0 certifications on their website. Two fast fashion brands, Fashion Nova and PrettyLittleThing, had 0 certifications on their websites. Refer to Table 4.2 for a breakdown of certifications found on each website. The most frequently occurring certifications were LEED and SGS, which were found on 4 of 10 websites. GOTS and SEDEX certifications were found on 3 websites, OEKO-TEX on 2 websites, and NAFTA and WRAP each on 1 website.

Brand	Certifications	Count
Forever21	Leadership in Energy and Environmental Design (LEED), Sociéte Générale de Surveillance (SGS)	2
SHEIN	Sociéte Générale de Surveillance (SGS)	1
Fashion Nova	-	0
H&M	Global Organic Textile Standard (GOTS), Leadership in Energy and Environmental Design (LEED), Société Générale de Surveillance (SGS), Supplier Ethical Data Exchange (SEDEX)	4
PrettyLittleThing	-	0
Big Bud Press	Global Organic Textile Standard (GOTS), North American Free Trade Agreement (NAFTA)	2
CHNGE	Global Organic Textile Standard (GOTS), Leadership in Energy and Environmental Design (LEED), OEKO-TEX	3
Fashion Brand Company	-	0
Tunnel Vision	Société Générale de Surveillance (SGS), Supplier Ethical Data Exchange (SEDEX), Worldwide Responsible Accredited Production (WRAP)	3
iGirl	Leadership in Energy and Environmental Design (LEED), OEKO-TEX, Supplier Ethical Data Exchange (SEDEX)	3

Table 4.2: Certifications found on each website.

After crawling the sitemap files for each website, it was found that fast fashion websites have a greater average number of pages compared to slow fashion websites. The fast fashion website page count ranged from 26,051 to 1,365,448 pages and the slow fashion website page count ranged from 398 to 21,453 pages. The average number of pages for fast fashion websites

was 345,706.4 pages and the average for slow fashion websites was 4,934.2 pages. Refer to Table 4.3 to view the page count for each website.

Brand	Estimated Number of Pages (Sitemap)
Forever21	30,619
SHEIN	1,365,448
Fashion Nova	275,326
H&M	26,051
PrettyLittleThing	31,088
Big Bud Press	961
CHNGE	770
Fashion Brand Company	398
Tunnel Vision	21,453
iGirl	1,089

Table 4.3: Estimated number of pages per website domain, counted via sitemaps.

The results show that the number of website categories is significantly higher in fast fashion brands compared to slow fashion brands. The number of website categories in the five fast fashion brands ranges from 318 to 1,019 categories. Meanwhile, the number of slow fashion website categories ranges from 16 to 53 categories. The average number of categories for fast fashion brands was 528 categories, while that for slow fashion was 30 categories. See Table 4.4 to view the number of categories for each brand's website.

Brand	Number of Website Categories
Forever21	318
SHEIN	413
Fashion Nova	1,019
H&M	401
PrettyLittleThing	489
Big Bud Press	53
CHNGE	32
Fashion Brand Company	19
Tunnel Vision	30
iGirl	16

Table 4.4: Number of categories per website domain.

The crawler results show that the number of cookies collected on retail websites are

similar on fast and slow fashion websites. The average number of cookies collected on fast fashion websites was 33 cookies, and on slow fashion websites the average was 25.8 cookies. Refer to Table 4.5 for the number of cookies counted on each website.

Brand	Number of Website Cookies
Forever21	27
SHEIN	45
Fashion Nova	49
H&M	17
PrettyLittleThing	27
Big Bud Press	19
CHNGE	35
Fashion Brand Company	24
Tunnel Vision	29
iGirl	22

Table 4.5: Number of cookies collected per website domain.

Through the research process, it was found that Fashion Nova, Big Bud Press, CHNGE, Fashion Brand Company, Tunnel Vision, and iGirl use Shopify for their website infrastructure. Shopify is a service that provides commerce support for businesses. All five slow fashion brands use Shopify, yet only one fast fashion brand (Fashion Nova) does. These six brands share similar website infrastructure, whereas the other four brands each have a unique and customized website structure.

Chapter 5

Discussion

While the differences in keywords, certifications, and cookies between fast and slow fashion websites were not significant, the results show that number of pages and categories are distinctly different. A larger number of pages corresponds to a fast fashion website, but this is more nuanced than a simple direct relationship. Tunnel Vision, a slow fashion brand, had over 21,000 pages on their website. This was closer to fast fashion brand H&M's 26,000 pages than any other slow fashion brand. Tunnel Vision, has thousands of listings by combining their house-brand designs with outside sourced vintage clothing. Despite their large number of products (and thus, pages), they qualify as a slow fashion company because their clothing is produced in small batches, they utilize sweatshop-free manufacturing, and they provide their employees with living wages [21]. Despite matching the website size of a fast fashion brand, Tunnel Vision remains categorized as slow fashion. How can a crawler grasp this context, when it is only built to capture individual data points?

If the number of pages is combined with the number of categories, this may add needed context. As mentioned, Tunnel Vision had over 21,000 pages, bringing them close to the number of pages on H&M's website. However, their website only had 30 categories, which is much lower than H&M's 401 categories. Fast fashion brands have a higher average number of categories because they have a vast range of products, including clothing, accessories, home, beauty, and more. They also include sale sections like "Under $5", "Under $10", and trending pages for new arrivals that have just been released. In contrast, the slow fashion

brands examined in this research have a more limited range of products, typically including just clothing and accessories, and they do not include more than one sale category or trending products.

If the data collected by this project's crawler is expanded to include other factors, such as price and size offerings, this would help to paint the bigger picture. While there are patterns in the fast and slow fashion website structure and content, all factors must be considered before determining a brand's type. This is especially important because information about a brand's manufacturer, labor practices, and materials sourcing must be considered. If a crawler can collect this data, then a machine learning model can be tasked with the classification of a website as fast or slow fashion. It can be trained to recognize which factors should be weighted more in the classification process (i.e., certifications, number of pages) as opposed to other factors (i.e., keywords).

For further research, the crawler can be improved to include more data points. As mentioned previously, it would greatly improve the classification process if sizes and prices were collected. One of the major proponents of fast fashion is its offering of trendy clothes at low prices. Slow fashion brands are often more expensive due to their quality, fabric sources, and labor practices [11].

While this research focused more on website content than structure, it would be beneficial to take a closer look at the structure of fast and slow fashion websites. It was found that all slow fashion brands utilize Shopify, while only 1 fast fashion brand, Fashion Nova, does so. Shopify provides administrative tools and a website-building service, along with running servers, software updates, security, and more [20]. Shopify is often recommended for smaller businesses as a "no-code" or "low-code" tool [8], which explains why all five slow fashion brands use it for their website and sales infrastructure. However, as shown with Fashion Nova, Shopify is not limited to small businesses and thus cannot be considered on its own when classifying fast and slow fashion brands.

Additionally, this research focused on crawling domains within the United States. To

increase the data collection, it can be expanded to other regions for comparison. This can show a difference in product offerings, which may be common in fast fashion companies that are larger and have more global reach. If the prices of products are collected globally, it can be used to show evidence of lower prices in lower-income countries and higher prices in higher-income countries, despite the product being the same [1]. Another opportunity for comparison exists in determining which countries can shop from various brands; some small businesses, such as select slow fashion brands, may be limited in where they ship their products.

Chapter 6

Conclusion

There are stark visual differences in the websites of fast and slow fashion brands, and this research has shined a light on some of the structural and content-based differences. Slow fashion brands are sustainability-focused, with transparency about their supply chain, labor sources, and manufacturing processes. Slow fashion brands are often small businesses and use their websites to communicate a brand story or mission, and their websites are often characterized by fewer pages and categories. On the other hand, fast fashion brands typically display a larger product range with lower prices, which lead to a greater number of pages and categories. These website differences are directly related to the business practices held by a brand. It is also important to note that these differences can vary, and some fast fashion brands may be taking steps towards sustainability or incorporating elements of slow fashion in their business practices and websites. Likewise, not all slow fashion brands will have the exact same approach to their online presence.

Of course, there may be factors not easily noticeable to the human eye. It is important not to fall victim to greenwashing or be misled by a brand that claims to be sustainable while maintaining the business practices of a fast fashion company. If fast and slow fashion brands can be determined solely from their website content and structure, this can greatly benefit the consumer experience for those trying to shift towards sustainably-sourced fashion.

Appendix A

Technology Reference

For further exploration of the technologies mentioned, visit the websites listed below.

Github Repository (Codebase):

- github.com/gracebmanning/gm-research-crawler-v2

Crawlers:

- Scrapy: scrapy.org
- Selenium: selenium.dev
- Heritrix: github.com/internetarchive/heritrix3/wiki
- Headless Chrome Crawler: github.com/yujiosaka/headless-chrome-crawler
- Puppeteer: pptr.dev
- Puppeteer Extra Plugin Stealth: npmjs.com/package/puppeteer-extra-plugin-stealth

Databases:

- Oracle XE: oracle.com/database/technologies/appdev/xe.html
- PostgreSQL: postgresql.org
- Redis: redis.io

Programming languages:

- Javascript: developer.mozilla.org/en-US/docs/Web/JavaScript
- jQuery: jquery.com

- Typescript: typescriptlang.org

Website Security:

- CAPTCHA: captcha.net
- Cloudflare: cloudflare.com

Supporting Technologies:

- Node.js: nodejs.org/en/about
- npm: npmjs.com/about
- GitHub: github.com/about
- Windows Subsystem for Linux (WSL): learn.microsoft.com/en-us/windows/wsl/about
- Amazon EC2: aws.amazon.com/ec2

Bibliography

[1] Adjusting for Price Differences Across the World, 10 2018.

[2] Fashion E-Commerce Global Market Report 2021: COVID-19 Growth and Change to 2030. Technical report, The Business Research Company, 5 2021.

[3] Donald E Eastlake 3rd and Paul Jones. US Secure Hash Algorithm 1 (SHA1). RFC 3174, 9 2001.

[4] Karen Becker-Olsen and Sean Potucek. Greenwashing. In S.O. Idowu, N. Capaldi, L. Zu, and A.D. Gupta, editors, *Encyclopedia of Corporate Social Responsibility*, pages 1318–1323. Springer, Berlin, Heidelberg, 2013.

[5] Allegra Catelli. How Fast, Cheap Fashion Is Polluting the Planet, 6 2023.

[6] Selcuk Cebi. Determining Importance Degrees of Website Design Parameters Based on Interactions and Types of Websites. *Decision Support Systems*, 54(2):1030–1043, 2013.

[7] Elise Dopson and Aaron Orendorff. The State of the Ecommerce Fashion Industry: Statistics, Trends & Strategies to Use in 2022, 2 2022.

[8] Gary Dushnitsky and Bryan K. Stroube. Low-code Entrepreneurship: Shopify and the Alternative Path to Growth. *Journal of Business Venturing Insights*, 16, 11 2021.

[9] Patrizia Gazzola, Enrica Pavione, Roberta Pezzetti, and Daniele Grechi. Trends in the Fashion Industry. The Perception of Sustainability and Circular Economy: A Gender/Generation Quantitative Approach. *Sustainability*, 12(7):2809, 2020.

[10] Pankaj Ghemawat and Jose Luis Nueno Iniesta. ZARA: Fast Fashion, 4 2003.

[11] Hasina Khatib. What Goes Into the Pricing of Sustainable Fashion? *Vogue India*, 5 2021.

[12] Christine Huynh. 10 Fast Fashion Brands We Avoid At All Costs, 2 2023.

[13] Hyojin Jung, Qian Xiong, and Tetsuya Sato. *Feature of Official Website Shop of Fast Fashion Brands*. PhD thesis, Kyoto Institute of Technology, Kyoto, 2013.

[14] MDN Web Docs. User-Agent.

[15] Carol Xiaojuan Ou and Choon Ling Sia. Consumer Trust and Distrust: An Issue of Website Design. *International Journal of Human-Computer Studies*, 68(12):913–934, 2010.

[16] Jean Paoli, Eve Maler, Michael Sperberg-McQueen, Tim Bray, and François Yergeau. Extensible Markup Language (XML) 1.0 (Fifth Edition). Technical report, W3C, 11 2008.

[17] Solene Rauturier. The 27 Ethical and Sustainable Fashion Terms You Need to Know, 8 2022.

[18] Solene Rauturier. What Is Fast Fashion and Why Is It So Bad?, 4 2022.

[19] Pedro Manuel do Espírito Santo and Antonio Trigo. Influence of Website Design on E-Trust and Positive Word of Mouth Intentions in E-Commerce Fashion Websites. *European, Mediterranean, and Middle Eastern Conference on Information Systems*, pages 316–330, 2020.

[20] Shopify Staff. What Is Shopify and How Does It Work? (2023), 2023.

[21] Tunnel Vision Team. About Us.