ZADÁNÍ DIPLOMOVÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Hanzl** Jméno: **Petr** Osobní číslo: **420866**

Fakulta/ústav: **Fakulta informačních technologií**

Zadávající katedra/ústav:

Studijní program: **Informatika**

Studijní obor: **Znalostní inženýrství**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Detekce temných vzorů v českých internetových obchodech

Název diplomové práce anglicky:

Detection of Dark Patterns on Czech Webshops

Pokyny pro vypracování:

The goal of the thesis is to analyze content on selected Czech webshops in order to detect so called dark patterns.
1. Analyze and describe existing methods for dark patterns detection in the Czech Web environment as well as in the world.
2. Design a crawler to retrieve Czech Webshops content and identify relevant product pages.
3. Implement the crawler and a method for dark patterns detection on selected Webshops.
4. Evaluate and describe results of your method.

Seznam doporučené literatury:

Jméno a pracoviště vedoucí(ho) diplomové práce:

doc. Ing. Tomáš Vitvar, Ph.D., katedra softwarového inženýrství FIT

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **16.02.2021** Termín odevzdání diplomové práce: _____

Platnost zadání diplomové práce: _____

doc. Ing. Tomáš Vitvar, Ph.D. podpis vedoucí(ho) práce	podpis vedoucí(ho) ústavu/katedry	doc. RNDr. Ing. Marcel Jiřina, Ph.D. podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

. _____ _____

Datum převzetí zadání Podpis studenta

Czech Technical University in Prague

Faculty of Information Technology

Department of Software Engineering

Master's thesis

Detection of Dark Patterns on Czech Webshops

Bc. Petr Hanzl

Supervisor: doc. Ing. Tomáš Vitvar, Ph.D.

16th of May, 2019

Acknowledgements

I wish to express my sicere thanks to my supervisor doc. Ing. Tomáš Vitvar, Ph.D. for the continuos encouragement.

I also thank my whole family, especially my parents for the support and attention.

Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended. In accordance with Article 46(6) of the Act, I hereby grant a nonexclusive authorization (license) to utilize this thesis, including any and all computer programs incorporated therein or attached thereto and all corresponding documentation (hereinafter collectively referred to as the "Work"), to any and all persons that wish to utilize the Work. Such persons are entitled to use the Work in any way (including for-profit purposes) that does not detract from its value. This authorization is not limited in terms of time, location and quantity.

In Prague on 16th of May, 2019 .....

Czech Technical University in Prague

Faculty of Information Technology

Citation of this thesis

HANZL, Petr. *Detection of Dark Patterns on Czech Webshops.* Master's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2021. Available also from WWW: ⟨`https://github.com/Lznah/DarkPatterns`⟩.

Abstrakt

Klíčová slova Temné vzory, Výpočty zaměřené na člověka, Strojové učení, Hierarchické hnízdění, Webové obchody

Abstract

Keywords Dark patterns, Human Centered Computing, Machine Learning, Hierarchical clustering, Webshops

Contents

List of Tables

List of Figures

Introduction

Dark patterns[5][9][12][10] are ways of designing a user interface of websites, apps or any other computer system in a specific way to trick, confuse or coerce a user in doing unwanted actions like confirming to share more information than is needed to use the service, signing up for things that the user did not mean to, buying unwanted products and more.

Typically, when the user reads a website or uses an app, he does not read all the words and makes quick assumptions[5]. Dark patterns then trick the user by hiding information of unpleasant truth. The user also trusts in his experience that he has gained from using other websites or apps and expects specific actions to happen or not to happen by using a similar pattern in the user interface. The user is tricked here by excepting this user interface behaviour, but in reality, it does something more or less than what the user expects[12]. Dark patterns are not only able to take advantage of the user not paying enough attention. Another dark pattern uses psychological methods to make users feel bad and guilty for not doing what the dark pattern wants them to do[12].

Research into tricky user interface designs and deceptive practices has surprisingly much history, but it was neglected for many years. In 1999, Hanson and Kysar were the first who examined how companies abuse customers' cognitive limitations and profit from. The rapid growth of the Internet and e-commerce increased more serious discussions and analyses of this topic. The term Dark Pattern itself was introduced by user interface expert Harry Brignull in 2010 to

create a library of different types of dark patterns and to shame websites using them[6].

In March 2021, the state of California added new regulation that now bans dark patterns that prevent users from opting out of the sale of their personal data[3]. Therefore, the topic of dark patterns becomes more and more relevant.

In 2019, a group of scientist from Princeton University introduced an automated approach that enables experts to identify dark patterns used on websites at scale[10].

This thesis's primary goal is to build on top of their research to analyse the prevalence of dark patterns on Czech webshops, also described in the Princeton study[10]. This thesis focuses on product pages and product purchase flow only because these are the most promising pages, where all the buying happens. Several subgoals need to be done to fulfil the primary goal:

- An automated approach of gathering data from numerous Czech webshops at scale.
- These extracted data needs to be analysed in order to train a model that can detect dark patterns.
- Evaluate and describe results.

The thesis does not aim to study the prevalence of dynamic dark patterns that display transients values over time.

Chapter 1

State of the art

Most studies [5] [9] [4] in the field of dark patterns have only described known existing types of dark patterns. Also, literature often proposes different dark pattern taxonomies. To find these patterns, scholars did manual research, analyzing page by page.

In contrast to this approach, that requires a lot of manual work, there is a study from Princeton University [10] and it proposes a completely new taxonomy. Not only the researchers recategorized and made more accurate the currently known types from the literature, but they were able to find new types of dark patterns, thus they extended the literature about these new types.

Princeton researchers also notes that only textual information on webshops were analysed. Continues, that the set of found dark pattern is restricted in this manner [10].

To find these new types, reseachers focused on product pages of webshops, because as they say, these pages are the most promising to contain dark patterns at any level of purchase flow [10]. Princeton Researchers did a lot of work to find these dark patterns. They separated it into three steps as it can be seen in figure 1.1.

Corpus Creation is the first step, there are several programmes to get domain names of webshops. They gathered websites with the highest Alexa Rank via Alexa Rank API. Then, they used paid service Webshrinker to filter out only those websites that are webshops. The list of domain still contained non-English

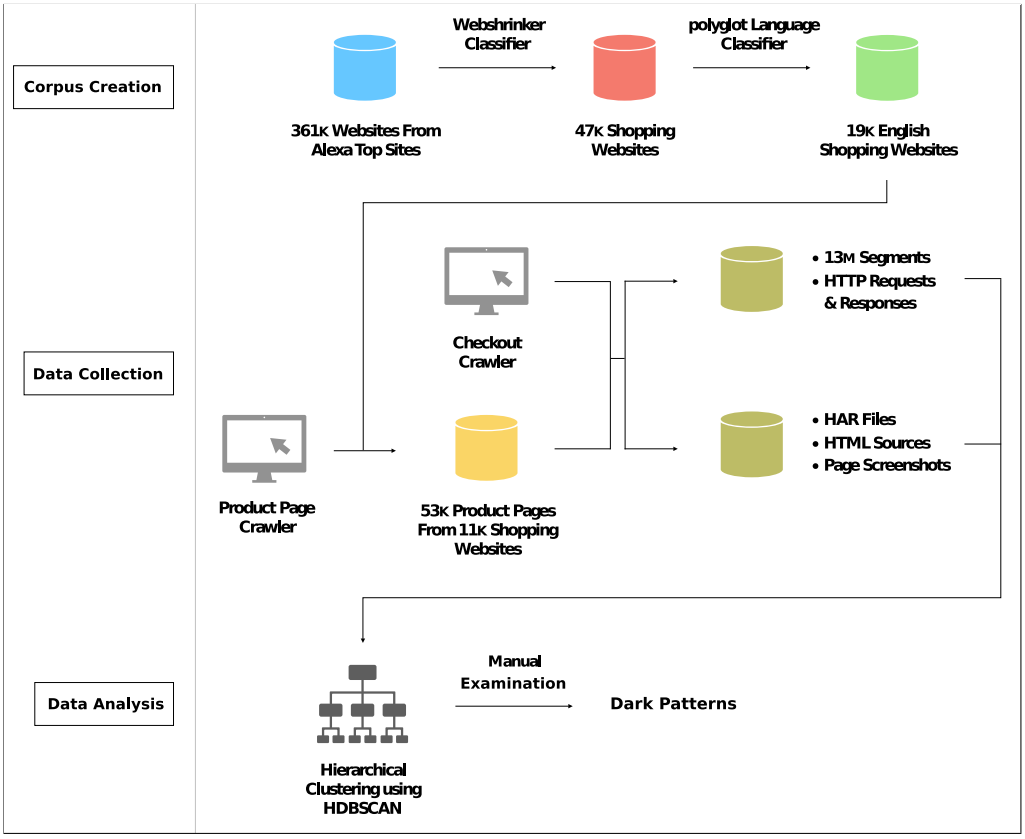

Figure 1.1: Overview of the shopping website corpus creation, data collection using crawling, and data analysis as proposed by Princeton University researchers[10].

websites. They used polyglot language classifier to filter them out of the list. Overall, researches gathered a list of 19K English shopping websites[10].

Data Collection is the second step. It consist of two crawlers created by the Princeton researches. The first crawler is meant to find product links on a single website. To speed up the process of finding these product pages, they trained a classifier of Logistic Regression on a dataset of 1000 URL links manually labeled by the researchers. The first crawler found 53K pages in 11K domain names.

The second crawler, also refered as 'checkout crawler', is meant to simulate user's shopping flow. This means that the crawler is able to follow the steps of the buying process - which includes selecting product options (e.g., size or color), adding the product to the card, viewing the cart and checking out. To

evaluate, whether or not this crawler is able to simulate user's shopping flow, the researchers randomly sampled 100 product pages and examined whether the crawler successfully reached the checkout page.

This crawler is build on OpenWPN, which is a web privacy measurement framework for privacy studies on a large set of websites[11]. Princeton researches implemented additional features to this framework. For example, they created a feature to store HAR files, which contain all the HTTP communication and Javascript calls. All these collected data are further utilized in an analysis phase by researchers. These data help researched to recognize whether or not a found pattern is on of the types of dark patterns.

The checkout crawler also divides visited pages into meaningful textual segments. Researchers propose the definition of this textual segment and an algorithm to split the HTML code of the page into these segments [10]. Also, the checkout crawler extracts data about text and background colors, positions and dimensions of the segments and others. With this algorithm, there were able to capture approximately 13 million segments across the previously noted 53K product URL pages.

Data Analysis is the last step of the research. It consists of data preprocessing, hierarchical clustering, examining and analyzing the found clusters. The data cleansing phase reduced 90% of all segments to 1.3 million segments.

Data were transformed into a representation of Bag of Words (BoW)[13]. Then, Principal Component Analysis was performed on the BoW matrix. The outcome were 3 components, which together represented 95% of the variance in the data.

Researchers chose an algorithms called Hierarchical Density-Based Spatial Clustering of Application with Noise (HDB-SCAN)[1] to find clusters in data. They tried different hyper parameters of this clustering algorithm and picked the most promising results.

Then, they did two passes in examining the clusters. In the first pass, they manually tagged clusters that can manifest as dark patterns. This pass reduced the number of the clusters from 10,277 to 1,768. In pass two, they manually

examined all the 1,768 clusters, whether the cluster contains any dark pattern [10].

Lastly, the researchers discussed the results and they iteratively grouped the discovered dark patterns into categories. They revealed 15 types of dark patterns in 7 categories on 1,254 websites, which represents 11,1% out of 10,277 [10].

Chapter 2

Dark Patterns

Dark Pattern is a relatively new term. This neologism was firstly used by Harry Brignull in 2010[8] when he registered a domain darkpatterns.org. In this domain, Brignull created an online library to share user interface patterns with deceptive characteristics that intentionally confuse and enrol users in unwanted situations. Another purpose of this online library is to shame websites that use dark patterns.

2.1 Definition

Brignull described dark patterns like so: 'Dark Patterns are tricks used in websites and apps that make you do things that you did not mean to, like buying or signing up for something.'[5] His definition is a simplification to quickly understand what dark patterns are. However, it does not include all the dark patterns that Brignull describes. For example, there is a dark pattern that purposely focuses users attention on doing one action and distracts their attention from alternatives.

A more accurate definition is the one used in the study made by Princeton researchers. They suggest this definition: 'Dark patterns are user interface design choices that benefit an online service by coercing, steering, or deceiving users into making decisions that, if fully informed and capable of selecting alternatives, they might not make.' [10]

2.2 Taxonomy

Brignull also defined the first types of dark patterns. This list of types is continuously updated when a new type of dark pattern is found. In April 2021, there were twelve different types of dark patterns defined[7].

The researchers from Princeton University have redefined this list considering the results of their study. This list consists of fifteen types of dark patterns and seven broad categories. Their work also differs from the prior work[5][2][4] by the new proposed taxonomy. This new taxonomy focuses on the characteristics of dark patterns and cognitive biases that they exploit in users. They used their taxonomy to classify and describe discovered dark patterns.

This thesis uses the same taxonomy defined by Princeton researchers. This taxonomy consists of five dimensions:

Asymmetric
: The user interface presents more alternatives to a user. It is an asymmetric characteristic of a dark pattern if the user interface requires less effort to continue with the alternative that might be disadvantageous for users. A typical example is buttons for accepting and rejecting cookies on websites. Usually, the rejecting button is less noticeable. Also, if users want to reject saving cookies, the user interface forces them to read much more text and click many buttons for every single cookie.

Covert
: The user interface shows evidence of covert characteristics if users may fail to recognise the intended outcome of a specific action. Users have experience with other user interfaces, and they may predict a similar outcome from the interface that shows similar traits as a decoy to influence their decision-making process. For instance, most of the websites offer a subscription to a newsletter in the process of registration. Usually, this subscription to the newsletter is done by ticking a checkbox in the registration form. When users start to read a sentence mentioning the subscription, they automatically expect that not ticking the checkbox means not subscribing to the newsletter.

Deceptive
: The user interface induces false beliefs in users by presenting them with

misleading information. For instance, a website may offer a discount for a limited period of time, but in reality, the discount is permanent. Another example is a website that shows how many users are watching the given product and how many products are in stock. This information can take advantage of the deal by steering users into making quick decisions or inducing false beliefs of the product's exclusivity.

Hides Information
: The user interface intentionally delay presenting necessary information in places or in time, where or when users do not expect them to be presented. For instance, a website may present extra fees for a bought product at the very last step of the checkout.

Restrictive
: The user interface restricts the set of choices available to users and takes advantage of it. For example, a website may require to sign up only with Facebook to collect additional personal information.

In addition to these dimensions, Princeton researchers define six different effects on users through exploiting different cognitive biases by specific dark patterns:

- **Anchoring Effect**: The tendency of users to over-rely on the first piece of information in the future decision-making process.
- **Bandwagon Effect**: The tendency of users to value more or believe in something simply because others do.
- **Default Effect**: The tendency of users to stick with default options.
- **Framing Effect**: The tendency of users to choose different options with knowledge of the same information, but with different way of presenting the options.
- **Scarcity Bias**: The tendency of users to value more things that are more sparce.
- **Sunk Cost Fallacy**: The tendency of users to continue an action, because they already invested time or other resources in it. Users tend to continue even if that action is capable to put them in an even worse situation.

2.3 Types of Dark Patterns

Types introduced in this section are the same, that are defined in the paper from Princeton university[10]. They are based on the types firstly published by Harry Brignull[5]. Princeton researchers discovered 15 types of dark patterns in total and they divided them into 7 broader categories. They are summariezed in the table ??.

Sneaking

It is an attempt to hide, disguise, or delay of information that is relevant to users. Users would likely change their action future action, if they knew about this information. There are 3 types of dark patterns in this category: Sneak into Basket, Hidden Costs, and Hidden Subscription. Examples of these dark patterns can be seen in figure ?? Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Sneak into Basket

This dark pattern adds aditional products into the user's basket without their consent. Usually, he is not aware of this fact. The added products are bonuses or additional services. For example additional year of warranty or a gift card. The important for this dark patterns is that it raises the total price and users migth not be aware of this fact. This dark pattern exploits the *default effect* cognitive bias in users that was described earlier in this thesis. Literature here says, that this dart pattern is not *covert*, because user can see the added products in their baskets.

Hidden Cost

This is an attempt to add additional charges, typically at the end of the purchage process. Typical examples of this type of dark pattern are additional service fees or handling costs. This type of dark pattern is also not *covert*, but it may be considered as partially *deceptive*, because the information is delayed from users.

Figure 2.1: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

Figure 2.2: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

Also, this dark pattern can be classified into *hides information* dimension, as it attempts to hide information from users.

Hidden Subscription

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Figure 2.3: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

Urgency

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Countdown Timer

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Limited-time Message

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between

Figure 2.4: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Misdirection

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Confirm shaming

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and

Figure 2.5: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

Figure 2.6: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Visual Interference

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Figure 2.7: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

Trick Questions

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Pressured Selling

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Social Proof

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some

Figure 2.8: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Activity Message

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Testimonials

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Figure 2.9: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

Figure 2.10: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

Scarcity

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Low-stock Message

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some

Figure 2.11: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

High-demand Message

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Obstruction

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Figure 2.12: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

Figure 2.13: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

Hard to Cancel

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Forced Action

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really?

Figure 2.14: One image, 60 % of line width. **[[Napsat pořádný titulek]]**

Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Forcel Enrollment

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Chapter 3

Methodology

Conclusion

Bibliography

1. BERBA, Pepe. *Understanding HDBSCAN and Density-Based Clustering* [online]. 2020 [visited on 2021-04-04]. Available from: https://towardsdatascience.com/understanding-hdbscan-and-density-based-clustering-121dbee1320e.
2. BÖSCH, Christoph; ERB, Benjamin; KARGL, Frank; KOPP, Henning; PFATTHEICHER, Stefan. Tales from the dark side: Privacy dark strategies and privacy dark patterns. *Proceedings on Privacy Enhancing Technologies*. 2016, vol. 2016, no. 4, pp. 237–254. Available from DOI: 10.1515/popets-2016-0038.
3. *California bans 'dark patterns' that trick users into giving away their personal data* [online]. 2021 [visited on 2021-03-30]. Available from: https://www.theverge.com/2021/3/16/22333506/california-bans-dark-patterns-opt-out-selling-data.
4. CONTI, Gregory; SOBIESK, Edward. Malicious Interface Design: Exploiting the User. In: *Proceedings of the 19th International Conference on World Wide Web*. Raleigh, North Carolina, USA: Association for Computing Machinery, 2010, pp. 271–280. WWW '10. ISBN 9781605587998. Available from DOI: 10.1145/1772690.1772719.
5. *Dark Patterns* [online]. 2010 [visited on 2021-03-28]. Available from: https://www.darkpatterns.org/.
6. *Dark Patterns - About us* [online]. 2010 [visited on 2021-04-01]. Available from: https://www.darkpatterns.org/about-us.

7. *Dark Patterns - Types of Dark Pattern* [online]. 2010 [visited on 2021-04-05]. Available from: https://www.darkpatterns.org/types-of-dark-pattern.

8. GRAUER, Yael. *Understanding HDBSCAN and Density-Based Clustering* [online]. 2016 [visited on 2021-04-05]. Available from: https://arstechnica.com/information-technology/2016/07/dark-patterns-are-designed-to-trick-you-and-theyre-all-over-the-web/.

9. GRAY, Colin M.; KOU, Yubo; BATTLES, Bryan; HOGGATT, Joseph; TOOMBS, Austin L. The Dark (Patterns) Side of UX Design. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–14. CHI '18. ISBN 9781450356206. Available from DOI: 10.1145/3173574.3174108.

10. MATHUR, Arunesh; ACAR, Gunes; FRIEDMAN, Michael; LUCHERINI, Elena; MAYER, Jonathan; CHETTY, Marshini; NARAYANAN, Arvind. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proc. ACM Hum.-Comput. Interact.* 2019, vol. 1, no. CSCW.

11. *OpenWPM* [online]. 2010 [visited on 2021-04-03]. Available from: https://github.com/mozilla/OpenWPM.

12. *The Year Dark Patterns Won* [online]. 2010 [visited on 2021-03-28]. Available from: https://www.fastcompany.com/3066586/the-year-dark-patterns-won.

13. ZHANG, Yin; JIN, Rong; ZHOU, Zhi-Hua. Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics*. 2010, vol. 1, pp. 43–52. Available from DOI: 10.1007/s13042-010-0001-0.

Appendix A

List of Acronyms

DP	Dark Pattern
JS	Javascript
HTTP	Hyper Text Transfer Protocol
HAR	HTTP Archive
HTML	Hyper Text Markup Language
CSS	Cascading Style Sheets
API	Application Programming Interface
HDB-SCAN	Hierarchical Density-Based Spatial Clustering of Application with Noise
BoW	Bag of Words
PCA	Principal Component Analysis

Appendix B

Supplemental Material

The source code of the thesis and the implementation can be found on the attached medium or online at GitHub.

Thesis `https://github.com/Lznah/master-thesis`

GraphEvolution `https://github.com/Lznah/DarkPatterns`

```
├── README.md ..... the file with a brief contents description
├── MT_Petr_Hanzl_2019.pdf ..... the thesis text in PDF format
├── DarkPatterns/ ..... repository for the prototype
│   └── src/ ..... source code of the prototype
└── thesis/ ..... the directory of LaTeX source codes of the thesis
```

Directory structure B.1: Contents of the attached medium