



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich



Tabita Arn

Quantifying Web Personalization

Semester Thesis SA-2015-01
Oktober 2014 to January 2015

Tutor: David Gugelmann
Supervisor: Bernhard Plattner

Abstract

Online sale is one of the fastest growing economic sectors. In 2012 the retail sale via internet in Switzerland achieved a revenue of 5.8 billion CHF [1]. More than half of the Swiss population uses online shops at least once a week, and the number is still growing [2]. Over the last years, the paradigm in the advertisement industry has changed. Instead of paying for ad space on websites with high quality content the advertisement providers nowadays buy user profiles that fit well to their campaign. Hence, the advertisements track the user while browsing the web.

In this thesis we evaluate personalized online advertisements. Our main focus is on analysing the web services involved when loading advertisements. In particular what the differences between personalized and non-personalized advertisements are. We find that personalized and non-personalized advertisements are loaded differently. Furthermore, we find that how the personalized ad is loaded only depends on the ad provider and not on the website embedding the ad. As a result, we see the same advertisements on different websites. That is, users are followed by ads while browsing on different websites. Further, we observe that the same ad space on the same website can be used by different ad providers.

Acknowledgements

I would like to thank Prof. Dr. Bernhard Plattner for giving me the opportunity to write my semester thesis in his research group.

Furthermore, I would like to thank my tutor David Gugelmann. His professional advice and support was very valuable to realize this thesis. I appreciate that his door was always open to discuss ideas and problems.

Tabita Arn
Zurich, Switzerland
January 2015

Contents

1	Introduction	7
1.1	Motivation	7
1.2	Aim	8
1.3	Overview	8
2	Real Time Advertisements	9
2.1	Creating User Profiles	9
2.1.1	Third-Party Tracking	9
2.1.2	Information Exchange Between Third Party Trackers	10
2.2	Displaying Real Time Advertisements	10
2.2.1	Real Time Bidding	10
3	Analysis of Swiss News Sites	13
3.1	Services Included by Websites	13
3.2	Embedded Advertisement Companies	17
4	Experiment Without Personalization	19
4.1	Test Setup	19
4.2	Data Processing	20
4.3	Results	20
5	Experiment with Personalization	23
5.1	Test Setup	23
5.2	Data Processing	24
5.3	Results	25
5.3.1	Process of Embedding Personalized Advertisements	27
5.3.2	Differences between Non-Personalized and Personalized Ads	30
6	Summary and Conclusion	33
7	Outlook	35
A	Additional Figures	37

Chapter 1

Introduction

1.1 Motivation

Online sale is one of the fastest growing economic sectors. In 2012 the retail sale via internet in Switzerland achieved a revenue of 5.8 billion CHF [1]. Already in 2008 37% of the Swiss population used the internet for purchasing goods on a daily basis [2]. 23% stated to be in online shops about once a week [2]. Hence, more than half of the Swiss population uses online shops at least once a week, and the number is still growing. Since online sale has a huge potential, advertisement companies and online shops are fighting for customers. Online ads play an important role in acquiring customers. Online advertisement companies as well as online shops are very much interested in showing the right ads to the right people at the right time. Thus, they try to personalize advertisements as much as possible.

Together with the personalization of advertisements, also comes a personalization of information. This phenomenon is called filter bubble [3]. Meaning that information displayed to one person may not be shown to another person, even if they enter the same URL. Which information is displayed or not is chosen based on information gathered about a user beforehand. This leads to users being isolated in their bubbles consisting of their interests, as shown in Figure 1.1. The distortion of information can also lead to wrong assumptions: Let us assume somebody would like to purchase a camera online. After checking the camera on the internet, ad networks realize that the user is interested in a particular camera and offer right this model in subsequently displayed ads. Even on web sites whose content is not related to cameras. The person might now assume, that this camera must be an outstanding camera, since it is displayed everywhere. However, the camera is only displayed because he looked for it. It does not give any information about the quality or popularity of the camera.

In this thesis we want to examine how, why and when personalized advertisements are shown.

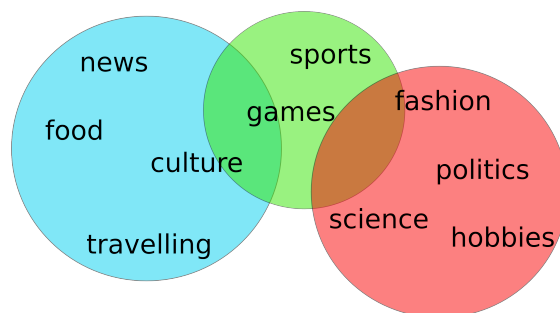


Figure 1.1: Visualization of the filter bubble phenomenon.

1.2 Aim

The objective of this project is to evaluate personalized online advertisements. We examine three questions:

- Which services are involved in displaying online advertisements?
- Which percentage of advertisements is personalized?
- Are there differences between personalized and non-personalized advertisements?

We conduct several experiments to answer these questions. First, we examine the most popular news sites according to Alexa [4]. We show which parties are involved in loading ads and how the advertisements are displayed. Second, we build personalized profiles on the top Swiss online shops according to revenue [5]. This allows us to learn which online shops do personalized advertisements and how. We then analyse the difference of personalized and non personalized online advertisements.

We focus on advertisements consisting of pictures and we do not investigate video or pure text advertisements to keep the scope of this thesis narrow.

1.3 Overview

Chapter 2 describes the concept of real time advertising. We discuss how information on a user is gathered and how the decision is made which advertisement to show to which user. Chapter 3 analyses the five most popular Swiss news sites. It highlights which parties are involved in displaying content in the user's browser and which of those parties are part of the online advertisement ecosystem.

In Chapter 4, we examine the advertisements displayed on news sites, if there is no browsing profile (information about the user's browsing behaviour) available to advertisement companies. In Chapter 5, we construct browsing profiles on online shops and we then again evaluate advertisements shown on news sites.

In Chapter 6 we summarize and discuss our results. We conclude with an outlook in Chapter 7.

Chapter 2

Real Time Advertisements

In this chapter we explain how the process of displaying personalized online advertisements in real time works. Section 2.1 enlightens how advertisement companies gain their knowledge about a potential customer. Section 2.2 explains the process of displaying real time advertisements (RTA).

2.1 Creating User Profiles

In order to be able to display personalized content on a website, the set-up of a user profile is crucial.

Ad companies gather information about users by tracking them over several websites. With the data they build a unique browsing profile for each user. This user profile enables them to display personalized content or perform, for example, site analytics. The displaying of targeted advertising is a very important ability for economical reasons. The price advertisement companies are ready to pay for being able to show a personalized ad is a lot higher, than for a non-personalized ad.

The collection of user data on the web is mostly done by third-party tracking.

2.1.1 Third-Party Tracking

This subsection is based on [6], which describes the different third-party tracking mechanisms on the web.

Third-party trackers are web services specialized on monitoring the activity of users on websites. Every website including the tracking code of a third party contributes to their user profile. If for example a user visits 20min.ch, his visit to 20min.ch is not only visible to 20min.ch. There are third-party trackers embedded into the 20min.ch web page such as doubleclick.net, which also log the user's visit. The third-party trackers are embedded in many different websites. This allows them to track the user over many websites and create a cross-site browsing profile of a user.

Let us now examine third-party tracking for targeted advertising, again using DoubleClick as an example. If a user visits a website where DoubleClick is in charge of advertising space, it will choose an advertisement to be embedded in the website. This ad is hosted by DoubleClick, therefore DoubleClick can set a cookie in the users browser belonging to DoubleClick. This cookie is always sent to DoubleClick whenever the user interacts with DoubleClick. DoubleClick can track the user over every website which embeds DoubleClick ads. This leads to a cross-site browsing profile.

2.1.2 Information Exchange Between Third Party Trackers

The different third-party tracking companies also exchange information about tracked users with each other [6]. There are complex business contracts and relationships in place. Therefore, we can not consider each tracking company in isolation. They do not only know what they observed themselves but also, what the other trackers shared with them.

If a tracker gets embedded in a website it can trigger further requests to other trackers by embedding objects hosted by these services. When embedding other tracker's objects, additional information, such as a unique user identifier, is usually provided as URL parameter. This enables the other trackers to also track the user. However, the trackers not embedded in the website relay entirely on information passed by the first tracker.[6]

2.2 Displaying Real Time Advertisements

With the evolution of the internet, there was also a new advertisement business established - online advertisement. At first the online advertisement companies focused on advertisement banners. However, this had the disadvantage that the advertisement companies had to buy expensive advertisement packages, which then were displayed to some customers interested in the product, but also to a lot of customers which had no intention to buy the product. With the raise of the search engines, the search engine marketing became popular. It offered more control of displaying advertisements to interested customers. Soon the capacity of search engine marketing was reached, since the demand was higher than the supply. Now, there is again a strong focus on the classic advertisement banners. This is due to the market saturation in engine marketing and the new possibility of personalized real time advertisement. The real time advertisement (RTA) allows the advertisement companies to decide in real time if they want to display an advertisement to a certain potential customer. Based on the knowledge about the customer, the decision is made. This allows to measure and optimize the return of investment.[7]

2.2.1 Real Time Bidding

Real time advertisement is based on real time bidding, the process of selling space in the browser window at auction in real time. The four main steps of real time bidding are as follows [7]:

1. A user visits a publishers [8] web page (e.g. 20min.ch), which sells allocated spaces for advertisements. An ad exchange (e.g. DoubleClick) enables the transaction between the publisher (ad seller) and the bidders (ad buyer) [8]. The ad exchange sends a request to the integrated bidders via their backend. The request contains different information, for example the advertisement space and a unique user ID. The user ID enables the bidders to verify if they already have a matching user profile.
2. The bidders verify whether the user profile data matches a target group of their advertisers. The advertisers are in the end paying for the advertisement and their ad is displayed. Mostly they do an auction within their advertisers. Based on this auction they send a bid for the advertisement space to the ad exchange.
3. The ad exchange assigns the advertisement space to the bidder with the highest bid.
4. The winning bidder displays his advertisement.

Since the loading of the advertisements must keep up with the loading of the other web page content, the assignment of the advertisement space must be done in real time. Hence, the whole process of selling the advertisement space and displaying the advertisement takes only up to around 100 milliseconds [7].

Figure 2.1 gives an overview about the relations between ad network and the involved parties.

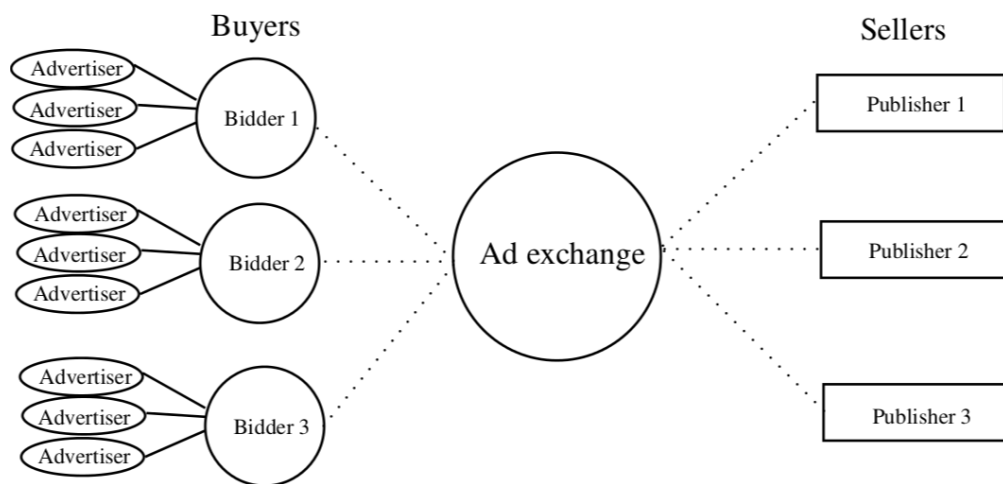


Figure 2.1: Structure of an ad network (from [8]).

Chapter 3

Analysis of Swiss News Sites

In this chapter we examine the top five Swiss news sites, according to Alexa [4] as of October 2014.

- Yahoo.com
- Blick.ch
- 20min.ch
- Bluewin.ch
- Tagesanzeiger.ch

We investigate which parties are involved in the process of loading the news site and we extract the parties responsible for displaying the advertisements.

3.1 Services Included by Websites

There is content delivered by many different sources. An example is shown in Figure 3.1. It visualizes the loading of Blick.ch. At first the user enters an URL (here Blick.ch) in the browser. In the background there are then many other URLs called, like e.g. Doubleclick.net. In the end all the gathered content is put together and shown in the browser.

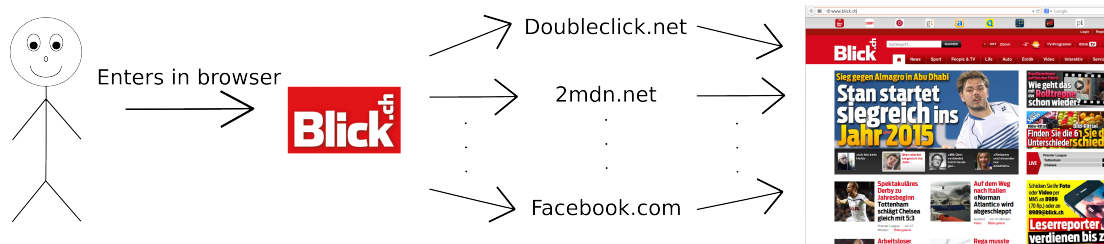


Figure 3.1: Loading of content from different domains for displaying Blick.ch in the browser.

Figures 3.2 to 3.6 show the involved Domains when loading the start page of 20min.ch, Blick.ch, Bluewin.ch, Tagesanzeiger.ch and Yahoo.com, respectively. The figures show, that many more than just one domain are involved in loading the page. 20min.ch has the most involved domains, in total 69 domains. On the other hand, yahoo.com calls only 13 different domains to load the content of its page. The exact number of domains can be found in Table 3.1. Table 3.2 contains the number of loaded, pictures, images and applications per domain. On average there are 171 pictures, 54 texts and 32 application loaded.

In the next section, we extract the embedded advertisement companies and evaluate their function.

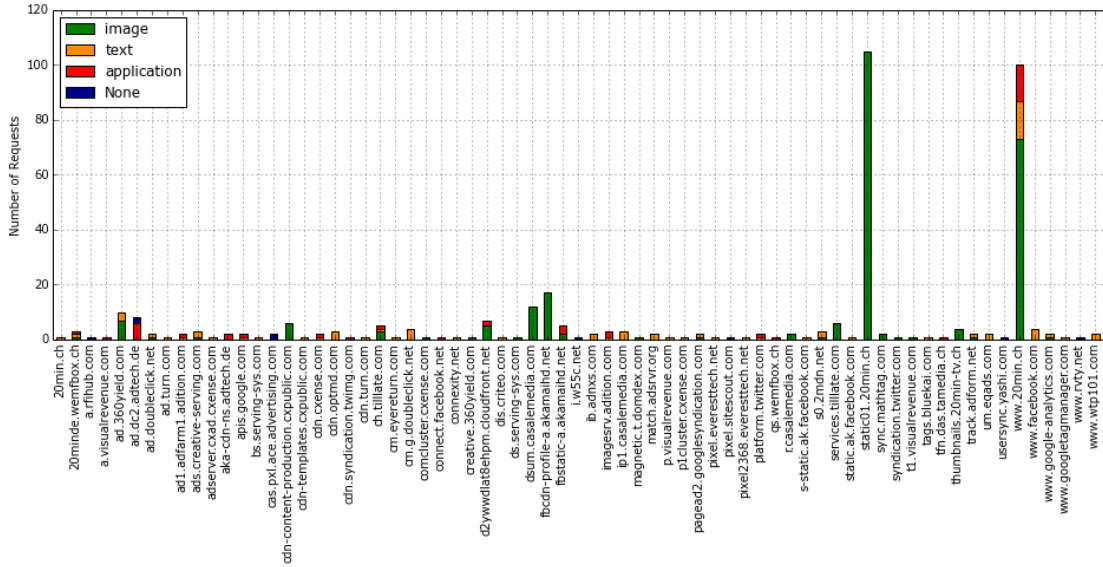


Figure 3.2: Domains involved in loading news site 20min.ch.

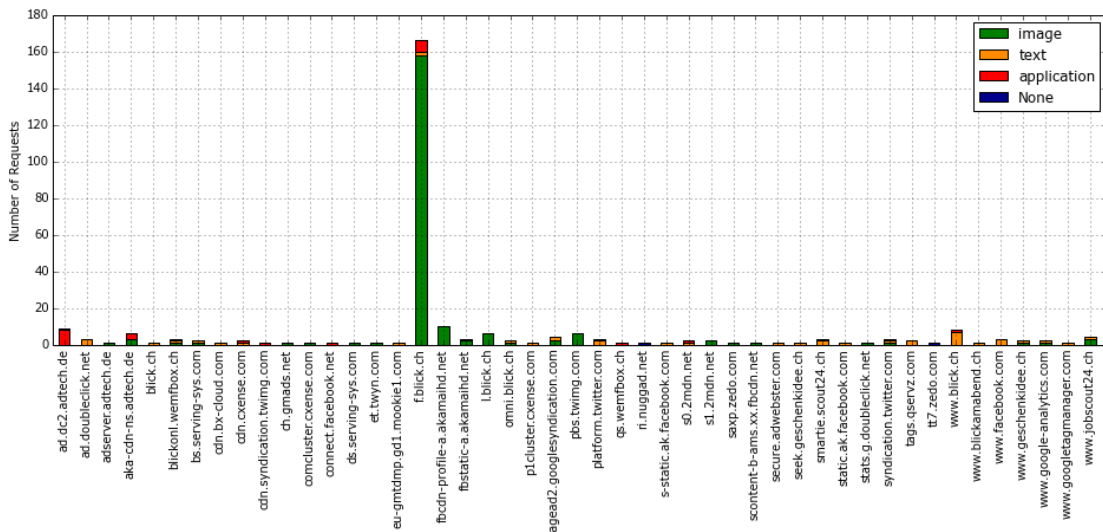


Figure 3.3: Domains involved in loading news site Blick.ch.

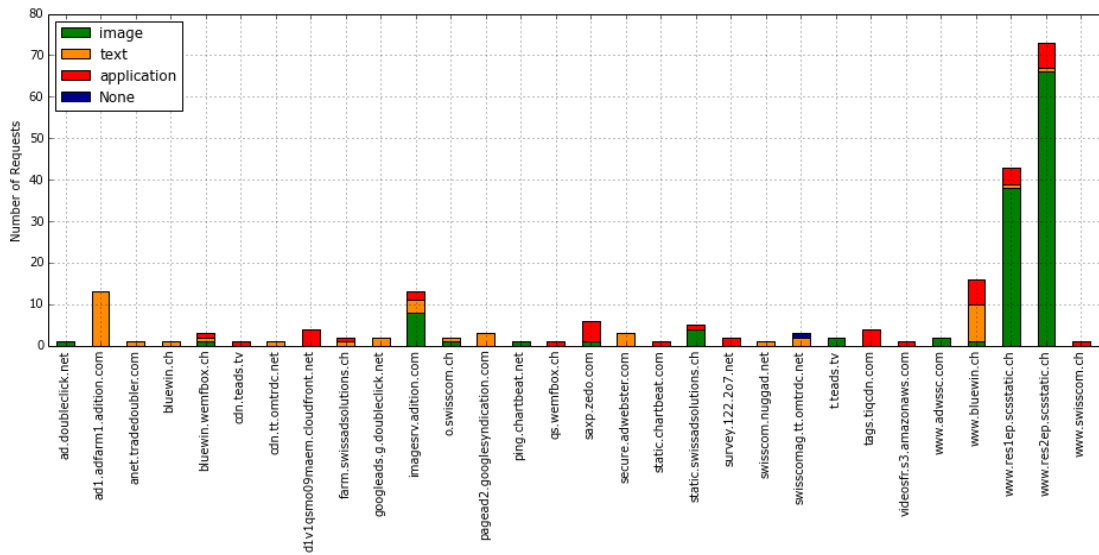


Figure 3.4: Domains involved in loading news site Bluewin.ch.

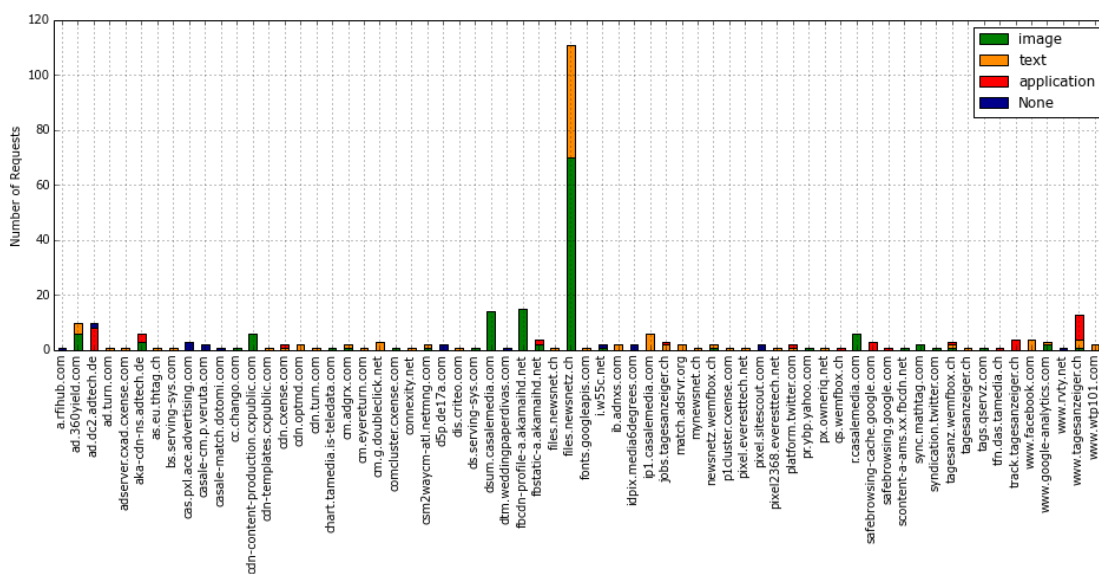


Figure 3.5: Domains involved in loading news site Tagesanzeiger.ch.

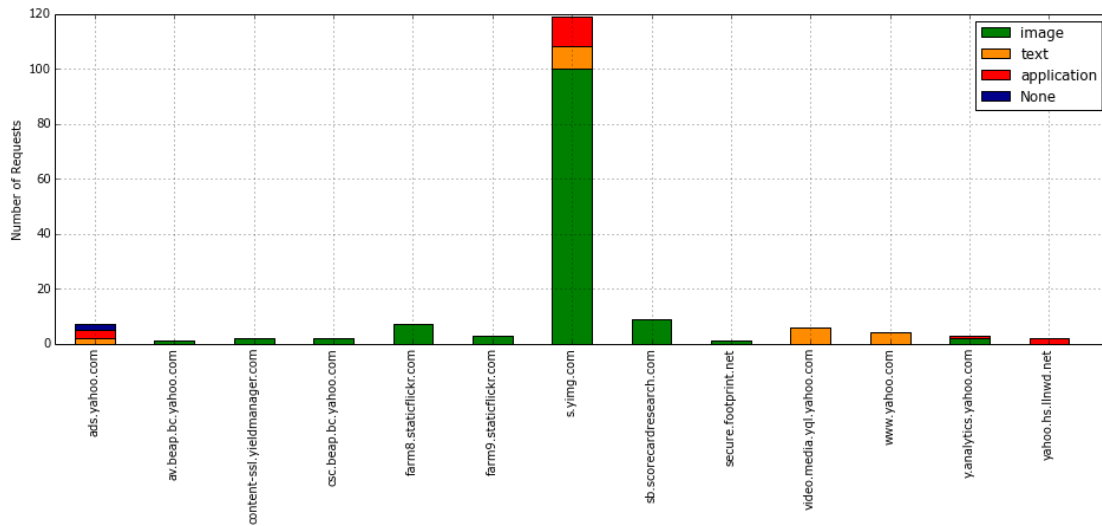


Figure 3.6: Domains involved in loading news site Yahoo.com.

Website	Number of Domains
20min	69
Blick	47
Bluewin	30
Tagesanzeiger	67
Yahoo	13

Table 3.1: Number of involved domains when loading landing page.

Website	Image	Text	Application	No Type
20min	257	71	39	9
Blick	206	42	28	3
Bluewin	126	44	41	1
Tagesanzeiger	140	95	36	18
Yahoo	127	20	17	2

Table 3.2: Content types when loading landing page.

3.2 Embedded Advertisement Companies

Displaying personalized online ads is a complex process. On one hand an ad service needs to have the browsing profile of a user in advance, which enables you to pick the right advertisement. This decision has to be done in real time which also adds up in complexity. The building of the browsing profile (see Chapter 2.1) and the decision making in real time (see Chapter 2.2) takes a lot of know-how and is only possible if you are able to track the user over many websites. Most companies who want to acquire customers with personalized ads do not have the necessary technical know-how and abilities. Thus, there is a big market for advertising companies. Those companies can have different roles, e.g. networks for tracking of users, algorithms on how to pick the right ads if given the browsing profile, composition of online ads or browsing statistics.

In this section, we examine the domains related to advertising in the five news sites: 20min.ch, Blick.ch, Bluewin.ch, Tagesanzeiger.ch and Yahoo.com. Whereby we focus on advertisement domains which do only advertisements, i.e. Facebook.com or Twitter.com are not listed as ad domains and are not further investigated.

Table 3.3 shows how many domains are involved in the corresponding news site and the percentage of second-level domains addressing online advertisement, whereby this can consist of analytics (tracking) or advertisement. For classifying the domains, we used the classified domains from [9] and the information stated on the domains homepage. If the second-level domain states on its website to be part of the online advertisement ecosystem or is listed as an ad or analytics domain in [9], we list it as an advertisement domain. However, if there is no information available, we do not list it as an ad domain. Therefore there might be even a few more ad domains. Hence, the numbers shown in Table 3.3 give a lower bound for the embedded advertisement domains.

On all news sites clearly more than 50% of all embedded domains are contributing to online advertisement. The median for online advertisement domains present on the news site is 34. All those domains are able to track you on the news sites.

Website	Domains	Ad Domains	Ad Domains %
20min.ch	72	44	61%
Blick.ch	56	34	61%
Bluewin.ch	37	24	65%
Tagesanzeiger.ch	67	41	61%
Yahoo.com	50	34	68%

Table 3.3: Number of total second-level domains and second-level ad domains on five news sites.

There are several ad domains which are present on all five news sites:

- Doubleclick.net
- Adnxs.com
- Googlesyndication.com
- Addition.com
- 2mdn.net

Those domains are within the top global players in online advertisement. There are five domains which are present on 20min, Blick, Bluewin and Tagesanzeiger, but not on Yahoo. This maybe since yahoo has their own advertising platform, and hence does not need to embed those domains. The five domains are:

- Wemfbox.ch
- Serving-sys.com
- Teads.tv
- Qservz.com
- Logentries.com

Table 3.4 shows what part of the loaded content is originating from ad domains and we highlight which percentage of the request are going to ad domains. If there would be no advertisements we could save 5 to 55 percent of the traffic caused by delivered content, on the evaluated five news sites. Yahoo has their own advertisement platform. Hence, they deliver most of their ads themselves. Since we only consider second-level ad domains which do exclusively online advertisement, we did not list Yahoo.com as an ad domain. Thus, the lower bound for Yahoo.com of 5% is not very accurate and the real value can be expected to be a lot higher. For the other four news sites the lower bound is quite accurate, since they only show advertisements from domains not owned by themselves. Thus, the loaded content originated from advertisement domains is rather 16 to 55% (see Table 3.4).

Website	Content delivered by ad domains	Request to ad domains%
20min.ch	43%	42%
Blick.ch	55%	53%
Bluewin.ch	24%	44%
Tagesanzeiger.ch	16%	36%
Yahoo.com	5%	10%

Table 3.4: Percentage of loaded content due to advertisements and percentage of requests to ad domains.

Chapter 4

Experiment Without Personalization

As a first experiment we examined how the news sites (20min.ch, Blick.ch, Bluewin.ch, Tagesanzeiger.ch, Yahoo.com) behave if we have no personalization. That means, we clean all cookies of the web browser before visiting the corresponding news sites such that there is no browsing profile of the user available. First, we explain how we proceeded in this experiment in Section 4.1. In Section 4.2 we shortly state the used data processing tools and procedure. We then evaluate the experiments in Section 4.3.

4.1 Test Setup

We try to ensure that there is no browsing profile of the user available, we conducted several steps:

1. Delete browser cache
2. Delete browser history
3. Acquire new IP address

We proceeded with the experiment as follows:

1. Perform steps to ensure there is no browsing profile available (see above)
2. Visit 10 links on news site (e.g. on 20min.ch)
3. Perform steps to ensure there is no browsing profile (see above)
4. Visit same 10 links on same news site
5. Start at step 1. with new news site

During the experiment we run a proxy which records the traffic. As a proxy we used the mitm proxy [10]. During the whole experiment we disabled the caching function of the proxy and the browser. The command to disable the caching functions and run the proxy used is:

```
1 mitmdump --host -v --anticache \  
2 -w pers.mitm > pers.log \  
3 --setheader :~s:Cache-Control:no-store --setheader :~s:Expires:0
```

Figure 4.1 displays an exemplary screen shot of the mitm proxy output.

```

127.0.0.1 GET http://thumbnails.20min-tv.ch/server063/381266/frame-31-381266.jpg
Host: thumbnails.20min-tv.ch
User-Agent: Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:31.0) Gecko/20100101 Firefox/31.0
Accept: image/png,image/*;q=0.8,*/*;q=0.5
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Referer: http://www.20min.ch/wissen/
Connection: keep-alive

<< 200 OK 5.39kB

Date: Tue, 06 Jan 2015 10:33:10 GMT
Server: Apache/2.2.22 (Debian)
X-Powered-By: PHP/5.4.4-14+deb7u7
Etag: "5a002dc-1594-50138be4bfa80"
Content-Length: 5524
Content-Type: image/jpg
Cache-Control: no-store
Expires: 0

127.0.0.1 GET http://www.google-analytics.com/analytics.js
Host: www.google-analytics.com
User-Agent: Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:31.0) Gecko/20100101 Firefox/31.0
Accept: */*
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Referer: http://www.20min.ch/wissen/
Connection: keep-alive

<< 200 OK 10.9kB

Date: Tue, 06 Jan 2015 09:08:52 GMT
Last-Modified: Thu, 13 Nov 2014 21:10:00 GMT
X-Content-Type-Options: nosniff
Content-Type: text/javascript
Vary: Accept-Encoding
Content-Encoding: gzip
Server: Golfe2
Content-Length: 11158
Age: 5058
Alternate-Protocol: 80:quic,p=0.02
Cache-Control: no-store
Expires: 0

```

Figure 4.1: Screen shot of the mitm proxy output while loading 20min.com

4.2 Data Processing

For the data processing we use the programming languages Bash and Python. For creating statistics we use the Python data analysis library Pandas [11].

We save the mitm proxy output of each clicked or entered URL in the browser in a separate file. In this way we can easily compare the delivered content of one loaded page with the content of the same loaded page in the second round, e.g. `http://www.bluewin.ch/de/sport.html` in the first round and `http://www.bluewin.ch/de/sport.html` in the second round.

To parse the mitm proxy output there is a Python library, `libmproxy` [12], available which we used. It allows to directly access all header fields and the content of each http request, e.g. accessing the content, content type and request domain:

```

1 fr = flow.FlowReader(open(PATH_MITM_FILE))
2 for i in fr.stream():
3     content = i.response.content
4     content_type = i.response.get_content_type()
5     domain = i.request.host

```

4.3 Results

In this section we want to evaluate whether the same ads are displayed to different users for whom there is no browsing profile available.

For the evaluation we compare the content of the websites with the same URLs, e.g. the content of `http://www.bluewin.ch/` in the first round and the content of `http://www.bluewin.ch/` in the second round (see Section 4.1). For each entered URL in the browser we sort the loaded content according to its type and store the hash of the content. This allows us to compare for

each URL the content loaded in the first and second round by comparing the hashes. This then tells us how much percent of the delivered content was the same.

Thereby we observe, that the delivered content is not at all the same, even without browsing profile available for personalization of the content. Figure 4.2 shows two screenshots of the same URL without personalization. It can be seen that the advertisements are not personalized but still different. In Table 4.1 we outline the results. The table shows the percentage of same content for each content type and news site. On average only 65% of the delivered images, 37% of the displayed text and 50% of the applications are the same when loading the same URL twice. This also means that about half of the loaded content is different.

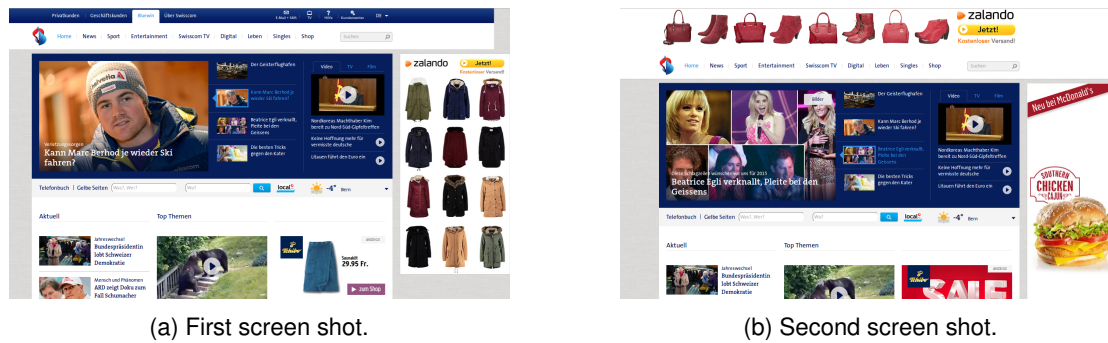


Figure 4.2: Two screen shots displaying <http://www.bluewin.ch/> without available browsing profile.

Website	Images	Text	Application %
20min.ch	61%	22%	41%
Blick.ch	56%	38%	32%
Bluewin.ch	74%	32%	60%
Tagesanzeiger.ch	76%	63%	50%
Yahoo.com	57%	28%	64%

Table 4.1: Percentage of same content on same URL for news sites and delivered content type.

We have shown that the news sites are not at all static. A large proportion of the content is dynamically allocated. The content of the news sites is constantly updated, reshaped and re-ordered. In addition, the advertisement are different when loading the same URL twice even without personalization. This does not depend on the personalization. They are changing with and without personalization. This fact does now impose a challenge to our next experiment, where we would like to examine personalized ads. Since it is far more difficult to show, that an advertisement is shown because of the personalized browsing profile and is thus only shown to the user with the according browsing profile and to nobody else, if the website is very dynamic. How we approached this challenge, is described in the next chapter 5.

Chapter 5

Experiment with Personalization

In our second experiment, we analyse personalized advertisements. To be able to analyse personalized advertisements, we first have to obtain the personalized ads. We describe how we retrieved these ads in Section 5.1. The data processing is explained in Section 5.2. We do the evaluation of the personalized advertisements in Section 5.3.

Please note, that we focus on advertisement banners, not on videos or advertisements consisting of text only.

5.1 Test Setup

We conduct the experiment to gain the personalized advertisements in two steps. First, we create a user with a browsing profile. In the second step we then evaluate on which news sites personalized advertisements matching the browsing profile are shown.

To build a user profile, we browse for 20 minutes on an online shop. Whereby we look for specific goods. In this way we established very distinct browsing profiles. We browsed one product type on one online shop for each browsing profile.

For selecting the online shops we used the online shops with the highest revenue in Switzerland according to carpathia [5]. We examined two selections. First, the five online shops with the highest revenue:

- Digitec.ch
- Amazon.de
- Nespresso.com
- Zalando.ch
- Leshop.ch

As a second choice, we focus on clothes online shops within the top 50 Swiss online shops, since we found during our pre-tests that clothing shops have an aggressive strategy with personalized advertisements. The evaluated online clothes shops are:

- Zalando.ch
- Laredoute.ch
- Bonprix.ch
- Jelmoli-shop.ch
- Fashionfriends.ch
- Cornelia.ch

- Manor.ch
- Espritshop.ch

After the browsing profile was created, we evaluated the ads on the five news sites 20min.ch, Blick.ch, Bluewin.ch, Tagesanzeiger.ch and Yahoo.com.

As we have seen in Section 4.3 the news sites are very dynamic. They display different advertisements even if there is no user profile available. How can we now verify if a advertisement is only shown to a user because of his browsing profile?

At each experiment we let two user browse in parallel, one with browsing profile and one without browsing profile. For the user without browsing profile we used an automatic browsing script SeleniumHQ [13] and for the user with browsing profile we conducted the browsing manually. We let the automatic script run on the five news sites for about six hours. In this way we gathered almost all the advertisement displayed on those news sites during this time frame. In parallel we manually browsed on the news sites. On each news site we performed ten clicks for each browsing profile. The traffic was again recorded by the mitm proxy [10].

In this way we could check which advertisements were only shown to the user with the browsing profile, but not to the automatic browser. To distinguish the personalized advertisements from the non-personalized advertisements, we took advantage of the personalized ads being only shown to the user with the according browsing profile.

In total we created 12 browsing profiles. The 12 browsing profiles and the goods we were looking for are shown in Table 5.1.

Online Shop	Good
Amazon.de	Climbing helmet
Bonprix.ch	Men, chequered shirt
Cornelia.ch	Men, black/grey jackets
Digitec.ch	Wireless phones
Espritshop.ch	Women, pattered skirts
Fashionfriends.ch	Women, bobble cap
Jelmoli-shop.ch	Men, short, colourful trousers
Laredoute.ch	Women, black dress
Leshop.ch	Swiss red vine
Manor.ch	Women, red blouse
Nespresso.com	Coffee machine
Zalando.ch	Men, colourful Nike trainers

Table 5.1: Description of browsing profiles created on the 12 online shops.

5.2 Data Processing

For the data processing we used the same tools as described in Section 4.2. However, we also needed further tools since we wanted to analyse the steps from entering the URL of the news site until the personalized advertisement was loaded and displayed. To be able to conduct this, we used NetworkX [14], a Python language software package for the study of complex networks.

In order to retain the sequence of http requests from entering the URL in the browser until the personalized ad was loaded we constructed a bidirectional NetworkX tree. The nodes in the tree represent the http requests. We then proceed as follows:

1. We load the mitm proxy output into a directed tree structure provided by networkX.

2. We search the node in the tree containing the domain delivering the personalized advertisement. This gives us the node containing the first request to the ad domain, the ad subtree root.
3. We travers the tree from the ad subtree root until the root node (entered URL in browser) and save the found path.
4. We cut the tree at the ad subtree root, resulting in two subtrees. This way, we obtain the ad subtree representing the loading of the ad.
5. We add the found path (step 3) to the ad subtree, see Figure 5.1.

This led to the complete tree representing the loading sequence of personalized online advertisements. Starting with the entering of the URL in the browser (root node) until the last requests for loading the advertisement are conducted (ad subtree leaves).

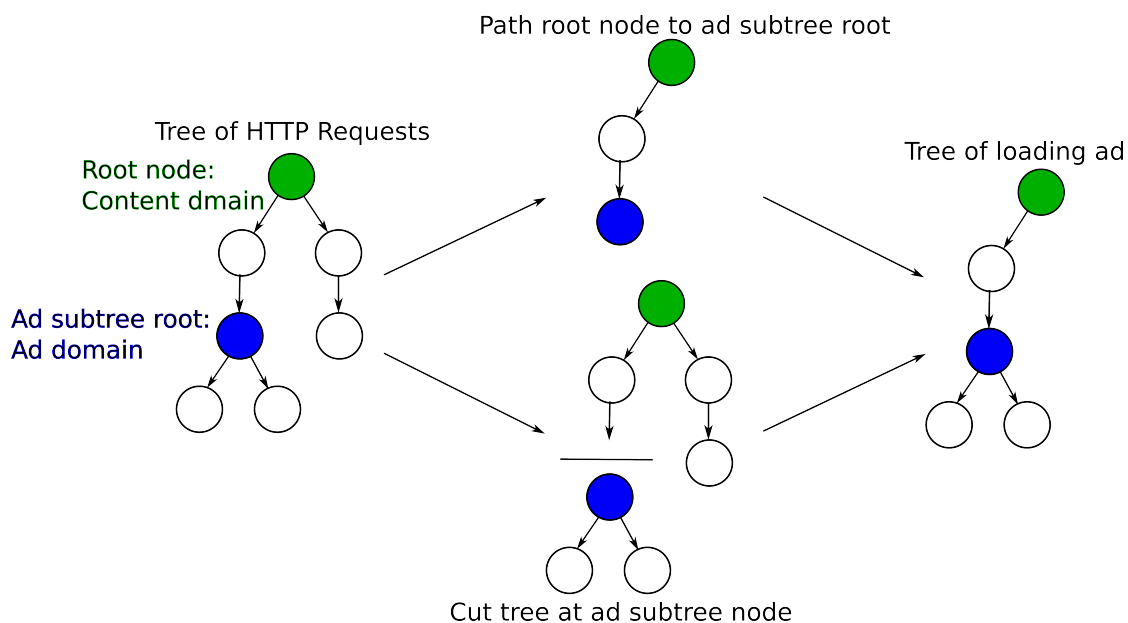


Figure 5.1: Constructing the tree representing loading personalized advertisements.

5.3 Results

From the first set of online shops (the five online shops with the highest revenue in Switzerland) only Zalando led to personalized advertisements on the news sites during our experiments. In the second selection (the online shops within the top 50 selling clothes) six out of eight browsing profile resulted in personalized ads. The browsing profile, created on the following online shops led to personalized advertisements:

- Zalando.ch
- Laredoute.ch
- Bonprix.ch
- Jelmoli-hop.ch
- Manor.ch
- Espritshop.ch

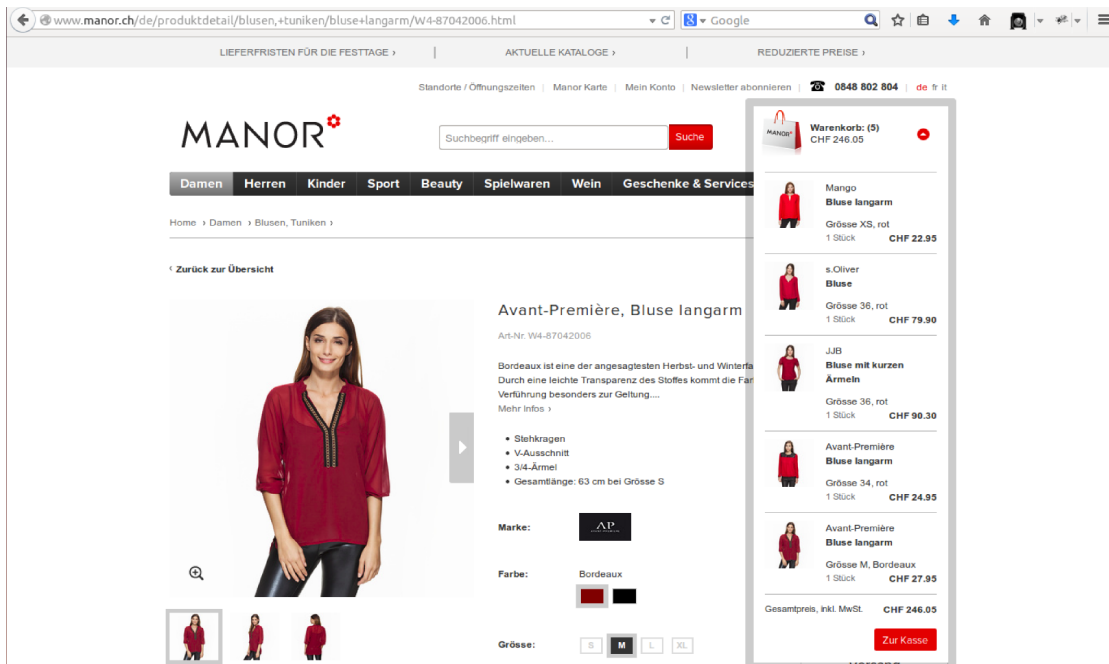
Figure 5.2 shows an exemplary shopping baskets and a screen shot of the according personalized advertisements. Screen shots of additional browsing profiles can be found in Appendix A. One can see, that the items in the shopping basket and the items displayed in the personalized advertisements are mostly the same. We can also see that not all the advertisements are

personalized. There is mostly personalized ads displayed, but also non-personalized ads.

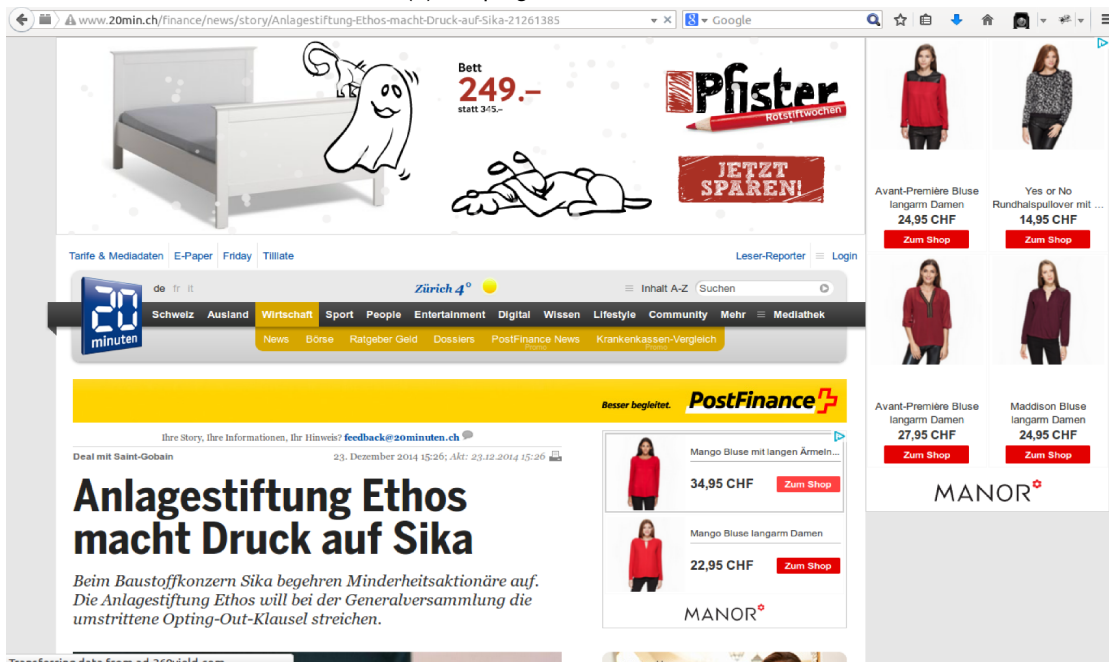
Table 5.2 shows which browsing profile led to personalized ads on which news sites. We examined, how the personalized ads are displayed and what the differences between the non-personalized ads and the personalized ads are. We summarize this in the following subsections.

Browsing Profile	20min	Blick	Bluewin	Tagesanzeiger	Yahoo
Bonprix.ch	✓	✗	✗	✓	✗
Espritshop.ch	✓	✗	✗	✓	✓
Jelmoli-shop.ch	✓	✗	✗	✓	✓
Laredoute.ch	✓	✗	✗	✓	✗
Manor.ch	✓	✗	✗	✓	✗
Zalando.ch	✗	✓	✗	✗	✗

Table 5.2: Shopping sites leading to personalized ads on different news sites.



(a) Shopping basket on Manor.



(b) Personalized advertisements shown on 20min.ch after adding items to the shopping basket on Manor.ch.

Figure 5.2: Screen shots of the shopping basket created on Manor.ch and the corresponding personalized advertisements shown on 20min.ch.

5.3.1 Process of Embedding Personalized Advertisements

Next we investigate how personalized advertisements are loaded on the different news sites. We find that the way the advertisement is loaded does not depend on the news site it is shown, but only on the advertisement provider in charge of displaying the advertisement.

In Figure 5.3 we visualise the platform independence of the personalized advertisements. The platform nodes can be exchanged and the nodes preceding the platform node do not change.

They only depend on which advertisement provider is doing the campaign.

Figures 5.4 to 5.10 show the tree of the http requests which lead to the loading of the personalized advertisement. We can see, that whenever Criteo is displaying the ad the tree looks the same, see Figures 5.5, 5.6, 5.7, 5.8 and 5.9. The trees for Yahoo, always call Yimg.com after the request to Yahoo.com. Yimg.com is owned by Yahoo and is used for their content delivery. Hence, it does not belong to the advertisement loading part, but is part of Yahoo's website structure.

We also observed that the same advertisements are shown on different platforms. This means that the advertisement provider tracks users over several platforms.

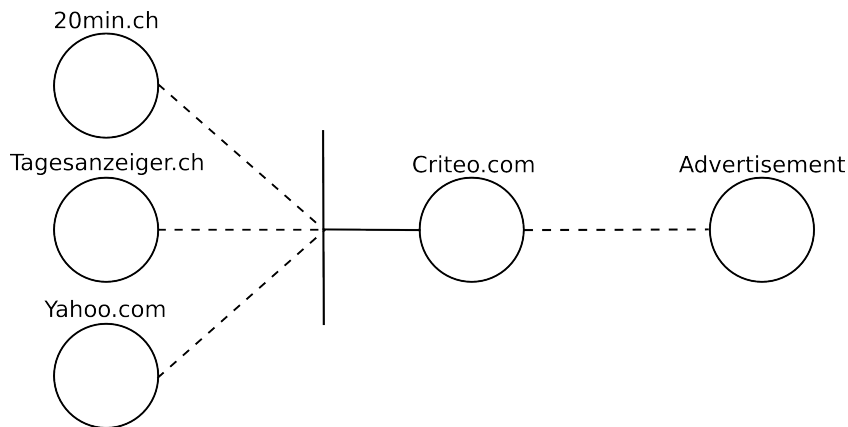


Figure 5.3: Visualization of Criteo's advertisements being platform independent.

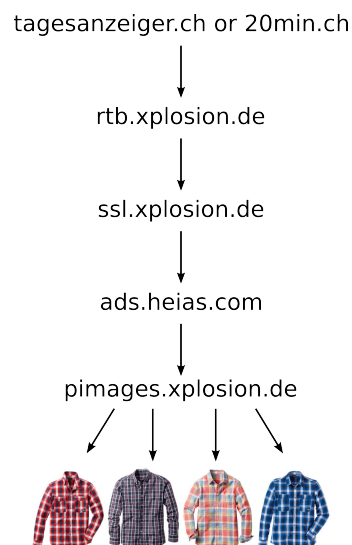


Figure 5.4: Requests for displaying personalized ads on news sites.

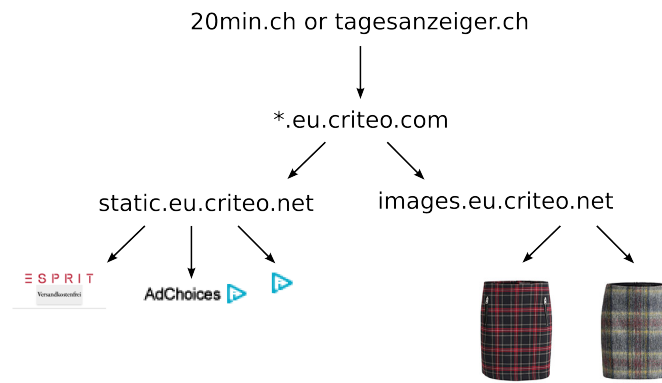


Figure 5.5: Requests for displaying personalized ads on news sites.

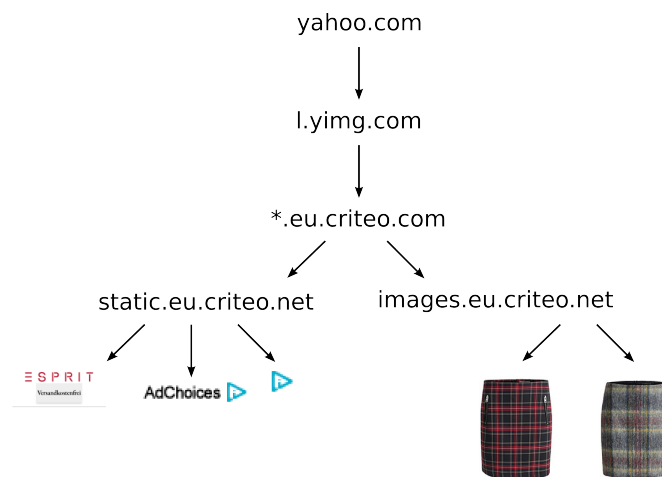


Figure 5.6: Requests for displaying personalized ads on Yahoo.com.

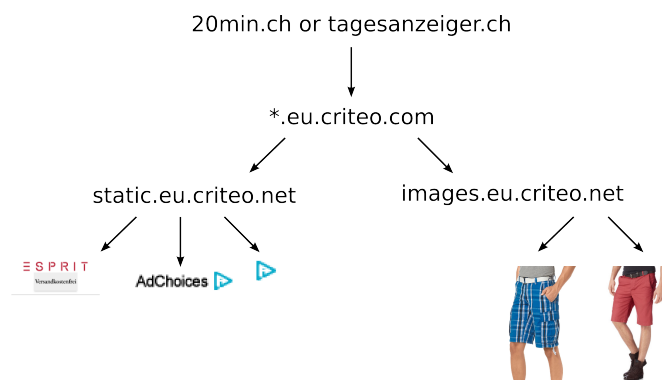


Figure 5.7: Requests for displaying personalized ads on news sites.

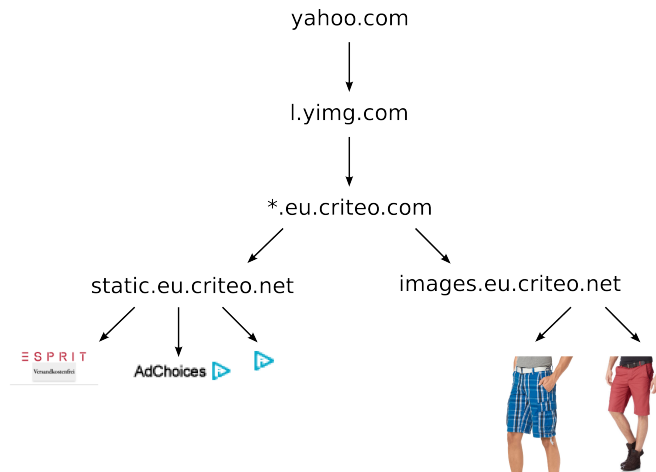


Figure 5.8: Requests for displaying personalized ads on Yahoo.com.

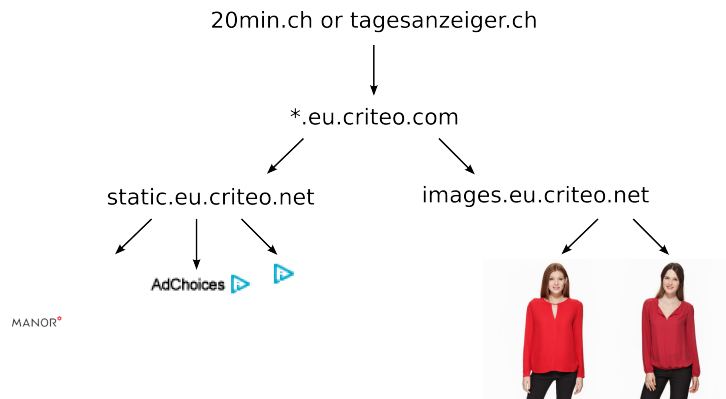


Figure 5.9: Requests for displaying personalized ads on news sites.

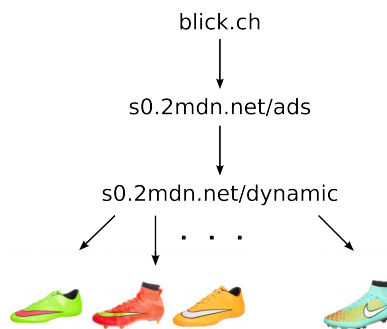


Figure 5.10: Requests for displaying personalized ads on Blick.ch.

5.3.2 Differences between Non-Personalized and Personalized Ads

The main difference between personalized and non-personalized advertisements is of course the content. The personalized advertisements show exactly the items we were looking at in the online shops plus a few matching recommendations. On the other hand, what is displayed in the non-personalized advertisements does not correspond to a website or item we were looking at.

The second difference, is to whom the advertisements are displayed. The non-personalized advertisements were also shown to the automatic browser without browsing profile. However, the personalized advertisements were exclusively shown to the user with the corresponding browsing profile.

Third, we found that the way how the advertisements are loaded is quite different. When loading a non-personalized advertisement the ad provider delivers the complete advertisement in one picture. However, if a personalized advertisement is loaded, the ad provider loads all the different images displayed in the advertisement separately. They are then put together directly in the users browser with the help of a JavaScript. This results in many request for only getting the content of one personalized advertisement. For example, Criteo loads about three static pictures (e.g. frame to put pictures of items with label of online shop) and it also loads up to eight pictures of the items (e.g. red blouses from Manor).

We observed that the news sites changes the ad providers quite frequently. The advertisement banners on the news sites change their ad provider even if there is no personalized advertisement shown. This may happen if you load a different URL on the same news site but also if you load the same URL again. We do not know how the mechanism behind this works. It could be that it is done in a random access manner or that there is a process run by the news site's server to chose the advertisement provider, which is not visible to the users.

Chapter 6

Summary and Conclusion

We found that to use online shops for creating a browsing profile works very well. This allowed to establish browsing profiles which led to personalized online advertisements in a short time. Furthermore, we found that online shops selling clothes are especially eager to show personalized online advertisements, so called retargeting [7].

We were able to elaborate different patterns on how personalized and non-personalized advertisements are loaded in the user's browser. The pattern for delivering non-personalized advertisements is as follows:

- There is one http call conducted to load the advertisement banner. The banner is delivered as one complete image.

For displaying the personalized advertisements this looks different:

- There are several http calls conducted to load the advertisement banner. The banner is delivered as many different images which are then put together with a Java script in the users browser.

This leads to many more requests for loading a personalized advertisement compared to a non-personalized advertisement.

Loading of a personalized advertisement does not depend on the website which embeds the advertisement. It only depends on which ad provider (e.g. Criteo, Xplosion) is in charge of loading the advertisement. Hence, if the personalized advertisement is displayed on different websites (e.g. 20min.ch and Tagesanzeiger.ch) and on both websites Criteo is in charge of embedding the ad, it is loaded in the same way.

The same personalized advertisements were displayed on different websites. This means, that the ad providers are able to track the users through several websites.

Furthermore, we observed that the same advertisement space on the same website can be used by different ad providers. If the same page is visited twice, the same banners may belong to different ad providers. We assume the advertisements might be allocated in a random manner or there could be a process running on the website's server to chose the advertisement provider based on certain criteria.

Another interesting finding of this thesis, is the fact that the evaluated news sites (20min.ch, Blick.ch, Bluewin.ch, Tagesanzeiger.ch and Yahoo.com) allocate their content in a very dynamic way. The content is continuously reordered, updated and exchanged. In order to be able to have such a dynamic website, a lot of different parties are involved in delivering content to the users browser. There are up to 69 different domains involved. On average around 30% of the traffic for loading one of those news sites is caused by companies belonging to the online advertisement ecosystem.

Chapter 7

Outlook

As a next step it would be interesting to create even more browsing profiles and evaluate the advertisements on several websites. This way, one could analyse personalized advertisements of additional advertisement providers.

It would also be interesting to create browsing profiles, which have been established over a longer time period. The assumption is, that the longer the advertisement companies know the user and therefore are more certain about the user being prone to buy certain items, the more willing they are to pay a higher price for being able to show the personalized advertisement [7].

Of course, it would be compelling to know the exact amount of money the advertisement providers pay for the advertisement space. However, as a user we only know who won the auction for ad space in the web browser, since we see the provider of the advertisement, but we do not know the exact amount paid. A way to circumvent this, would be to set up an own website which sells advertisement space. Hence, we would see how much money we made with which browsing profile, since we are the user and the website.

Appendix A

Additional Figures

Artikel	Verfügbarkeit	Anzahl	Gesamtpreis
Langarmhemd Art-Nr.: 90537081 Farbe: grau/rosa kariert Grösse: 39/40 (M) Preis: CHF 14.95	Lieferbar.	1	CHF 14.95
Langarmhemd Art-Nr.: 92128181 Farbe: hellblau/blau/dunkelorange/gelb Grösse: 41/42 Preis: CHF 12.95	Lieferbar.	1	CHF 12.95
Langarmhemd Art-Nr.: 94738681 Farbe: natur kariert Grösse: 41/42 (L) Preis: CHF 14.95	Lieferbar.	1	CHF 14.95
Langarmhemd Art-Nr.: 93325681 Farbe: sandrot kariert Grösse: 41/42 (L) Preis: CHF 14.95	Lieferbar.	1	CHF 14.95
Langarmhemd Art-Nr.: 94660795 Farbe: rot kariert Grösse: 39/40 (M) Preis: CHF 34.95	Lieferbar.	1	CHF 34.95

Artikelpositionen: 5
Bestellwert: CHF 92.75
zzgl. Versandkosten: CHF 6.50
Gesamtsumme: CHF 99.25

Algemeine Geschäftsbedingungen >
Datenschutz >

Sicherer Online-Kauf

Sichere Zahlung
Ihre Daten werden verschlüsselt und damit sicher übertragen.

Top Qualität
Wir prüfen permanent Material und Haltbarkeit unserer Produkte.

Sie brauchen Hilfe?

Rufen Sie uns an:
0848 - 100 150
0848 - 100 151
0848 - 100 152

Öffnungszeiten/ Horaires/
Orario di apertura

Schreiben Sie uns:
servicech@bonprix-shop.ch >

(a) Shopping basket Bon Prix.

Letztes Update: 10:59 Uhr

Tagesanzeiger

Zürich Schweiz International Wirtschaft Börse Sport Kultur Leben Wissen Auto Blogs Panorama Mehr -

Front

Meine neue Nummer 1 für mobile Telefonie.
Sunrise hat das beste Netz für mobile Telefonie.
connect SEHR GUT

Schlimmer als die Russen-Krise

Meistgelesen Newsticker

Transferring data from v4.moadads.com...

(b) Personalized ad Bon Prix.

Figure A.1: Screen shots of the shopping basket created on Bonprix.ch and the corresponding personalized advertisements on Tagesanzeiger.ch.

Artikel	Beschreibung	Einzelpreis	Anzahl	Gesamt
	Strukturierter Tallenrock 124CC1D006 Farbe: BLACK COLORWAY Grösse: 32	CHF 79,90	1	CHF 79,90
	Stretchiger Karo-Rock mit Zippertaschen 094EE1D016 Farbe: NIGHTBLUE Grösse: 38	CHF 79,90	1	CHF 79,90
	Stretchiger Karo-Rock in Woll-Optik 114EE1D017 Farbe: BLACK Grösse: 32	CHF 99,90	1	CHF 99,90
	Wolliger Karorock mit Alpakawolle 104EE1D013 Farbe: MULTICOLOUR Grösse: 32	CHF 109,90	1	CHF 109,90
	Leo-Print-Rock mit Tunnelzug 064EE1D030 Farbe: LEOPARD COLOURWAY Grösse: 32	CHF 65,90 CHF 49,95	1	CHF 49,95

Gutscheinodes können im Kassensbereich eingegeben werden.

(a) Shopping basket Esprit.

ESPRIT Leg lamp stolen from New York store's display

NORTH TONAWANDA, N.Y. (AP) — Who took the leg lamp belonging to Ralphie's "Old Man" from a New York store's annual tribute to "A Christmas Story"?

The owner of the Yankee Spirits liquor store in North Tonawanda says his annual window display honoring the holiday classic includes memorabilia from the 1983 film, including several versions of the now-iconic leg lamp.

The store's surveillance video shows a man in a gray hoodie and sunglasses walking into the store and glancing around before he grabs a large

ESPRIT Leg lamp stolen from New York store's display

Wolliger Karorock mit Alpakawolle
CHF 109,90

Stretchiger Karo-Rock in Woll-Optik
CHF 99,90

Stretchiger Karo-Rock mit Zippertaschen
CHF 79,90

Follow Yahoo News

Top Stories






N. Korea's Internet back online after outage

New York mayor calls for pause in protests after police ...


Argentine Orangutan unfazed by right to

(b) Personalized ad Esprit.

Figure A.2: Screen shots of the shopping basket created on Espritshop.ch and the corresponding personalized advertisements on Yahoo.com.

Artikel	Anzahl	Preis (CHF)
 XINT Chino Shorts regular fit türkis, M / ändern 483549F (1 Stk. = CHF 99.95) Nachlieferung, kommt in 3 Wochen	1	99.95 ×
 Levi's® Shorts Rot, 30, Länge 30 / ändern 73575791 (1 Stk. = CHF 129.90) sofort lieferbar	1	129.90 ×
 John Devin Bermudas Rosa-Kariert, 29, Normalgrößen / ändern 27529372 (1 Stk. = CHF 24.95) sofort lieferbar	1	24.95 ×
 John Devin Bermudas Gelb, 29, Normalgrößen / ändern 809784D (1 Stk. = CHF 26.95) sofort lieferbar	1	26.95 ×
 John Devin Bermudas blau-kariert, 29, Normalgrößen / ändern 79093491 (1 Stk. = CHF 59.90) sofort lieferbar	1	59.90 ×

Gutschein / Rabatt einlösen

 Rabattcodes können nach erfolgreichem Login auf Schritt „Bestellung prüfen“ eingegeben werden.

(a) Shopping basket Jelmoli.

de it it Zürich 6° Inhalt A-Z Suchen

Schweiz Ausland Wirtschaft Sport People Entertainment Digital Wissen Lifestyle Community Mehr Mediathek

News History Gesundheit Aus- und Weiterbildung Dossiers

Sofa **999.-** statt 1390.-

Pfister Rotstiftwochen bis 31.1.2015 **JETZT SPAREN!**

Ihre Story, Ihre Informationen, Ihr Hinweis? feedback@20minuten.ch

Medizinische Weisheiten 22. Dezember 2014 22:05; Akt: 22.12.2014 22:05

Können Sie Wahrheit von Dichtung unterscheiden?

Gut gemeint ist nicht immer auch gut – gerade punkto medizinischer Ratschläge. Testen Sie im Quiz, ob in Ihnen ein Mediziner oder eher ein Quacksalber steckt.

QUIZ: WAS IST MEDIZINISCHER FAKT UND WAS MYTHOS?

Strom mit Sparpotenzial: Was zu tun ist, wenn die Stromkosten in die Höhe steigen und was ein neuer Energieversorger Sinn macht, erfahren Sie in unserem Ratgeber.

Digitaler Unruhestifter: **Wer auf dem iPad liest, ist 1,5 Stunden länger wach**

Tablets sind leicht, handlich und bieten viel Speicherplatz. Aber:

(b) Personalized ad Jelmoli.

Figure A.3: Screen shots of the shopping basket created on Jelmoli-shop.ch and the corresponding personalized advertisements on 20min.ch.

Shopping basket LaRedoute

Item	Original Price	Discount	Final Price
Stretchkleid MINI PREIS	CHF 49.90	-	CHF 49.90
Kleid MINI PREIS	CHF 44.90	-30%	CHF 31.40
Etuikleid LAURA CLEMENT	CHF 89.90	-30%	CHF 62.90
Kleid mit Schmuckelementen SEE U SOON	CHF 119.00	-20%	CHF 95.20
Etuikleid, Materialmix LAURA CLEMENT	CHF 89.90	-	CHF 89.90

ihre BESTELLUNG

In Ihrem Warenkorb: CHF 393.60
 Ihre Reduktion: -CHF 64.30
 Versandkosten: GRATIS
GESAMTBETRAG: CHF 329.30

Bestellung abschliessen / **Einkauf fortsetzen**

(a) Shopping basket LaRedoute.

Personalized ad LaRedoute

LaRedoute **GRATIS LIEFERUNG***

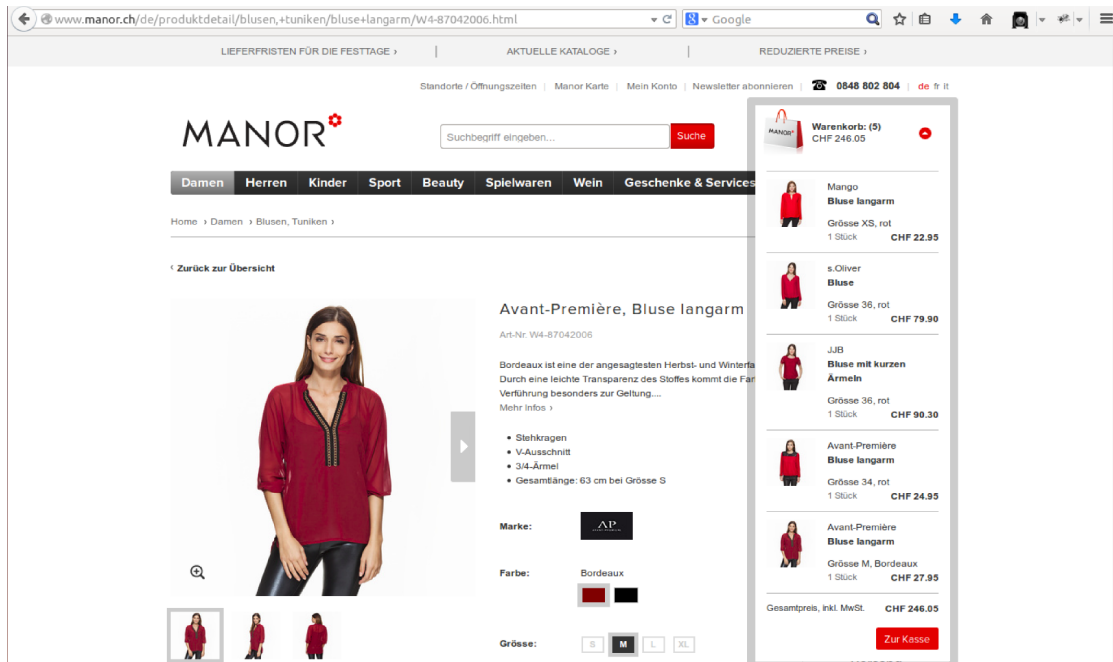
Stretchkleid	49,90 CHF	Zum Shop
Etuikleid, Materialmix	89,90 CHF	Zum Shop

Meistgelesen Newsticker

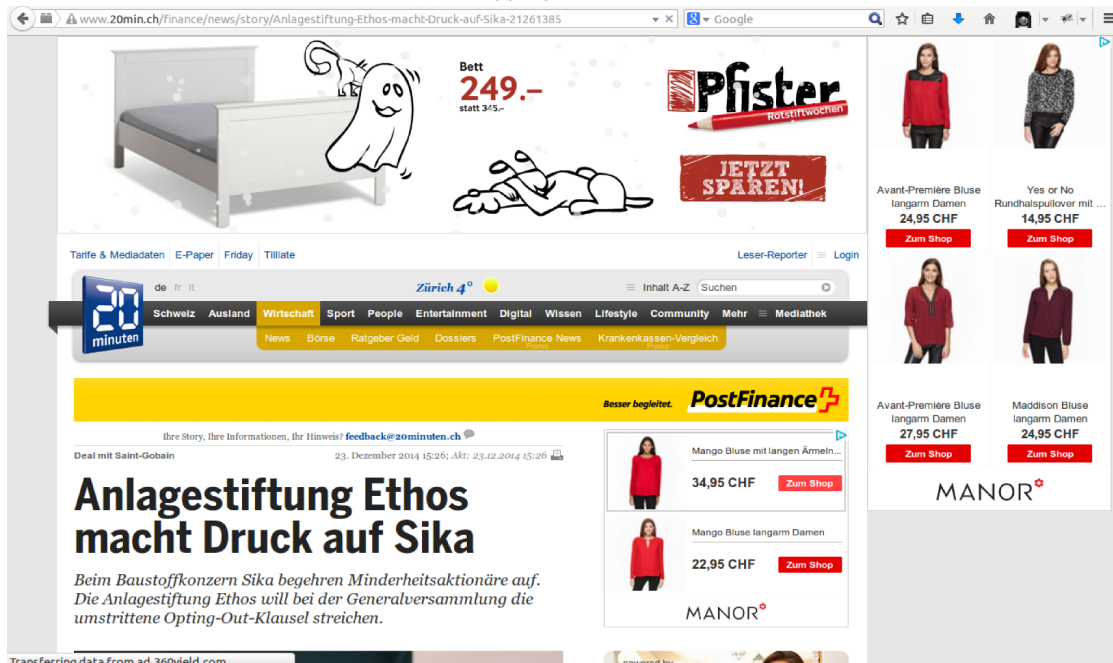
- Der SNB-Coup aus Sicht von Devisenhändler Bollier
- Kim offline - «Nordkorea ist restlos unten»
- Das Superschiff, das keiner haben will

(b) Personalized ad LaRedoute.

Figure A.4: Screen shots of the shopping basket created on Laredoute.ch and the corresponding personalized advertisements on Tagesanzeiger.ch.

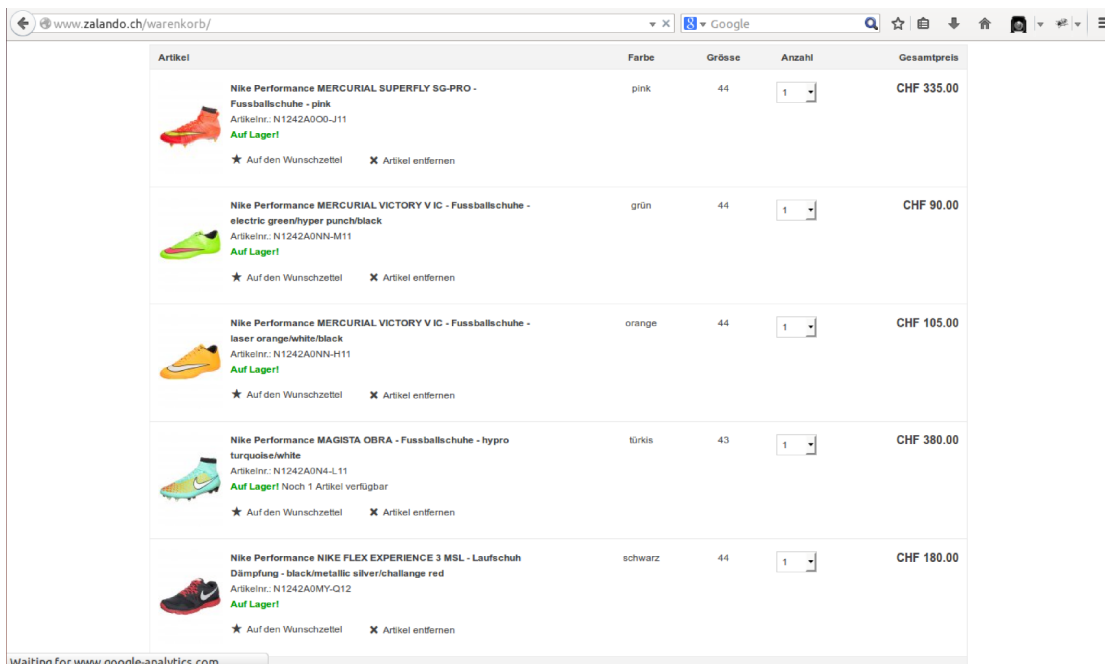


(a) Shopping basket Manor.

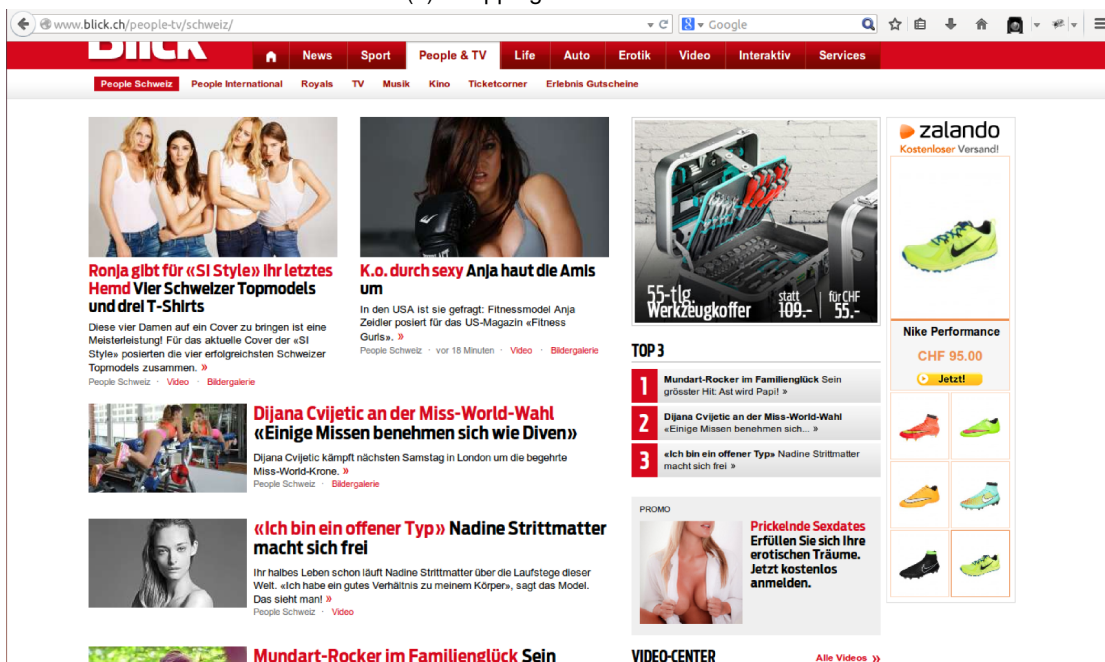


(b) Personalized ad Manor.

Figure A.5: Screen shots of the shopping basket created on Manor.ch and the corresponding personalized advertisements on 20min.ch.



(a) Shopping basket Zalando.



(b) Personalized ad Zalando.

Figure A.6: Screen shots of the shopping basket created on Zalando.ch and the corresponding personalized advertisements on Blick.ch.

List of Figures

1.1	Visualization of the filter bubble phenomenon.	7
2.1	Structure of an ad network (from [8]).	11
3.1	Loading of content from different domains for displaying Blick.ch in the browser.	13
3.2	Domains involved in loading news site 20min.ch.	14
3.3	Domains involved in loading news site Blick.ch.	14
3.4	Domains involved in loading news site Bluewin.ch.	15
3.5	Domains involved in loading news site Tagesanzeiger.ch.	15
3.6	Domains involved in loading news site Yahoo.com.	16
4.1	Screen shot of the mitm proxy output while loading 20min.com	20
4.2	Two screen shots displaying http://www.bluewin.ch/ without available browsing profile.	21
5.1	Constructing the tree representing loading personalized advertisements.	25
5.2	Screen shots of the shopping basket created on Manor.ch and the corresponding personalized advertisements shown on 20min.ch.	27
5.3	Visualization of Criteo's advertisements being platform independent.	28
5.4	Requests for displaying personalized ads on news sites.	28
5.5	Requests for displaying personalized ads on news sites.	29
5.6	Requests for displaying personalized ads on Yahoo.com.	29
5.7	Requests for displaying personalized ads on news sites.	29
5.8	Requests for displaying personalized ads on Yahoo.com.	30
5.9	Requests for displaying personalized ads on news sites.	30
5.10	Requests for displaying personalized ads on Blick.ch.	30
A.1	Screen shots of the shopping basket created on Bonprix.ch and the corresponding personalized advertisements on Tagesanzeiger.ch.	38
A.2	Screen shots of the shopping basket created on Espritshop.ch and the corresponding personalized advertisements on Yahoo.com.	39
A.3	Screen shots of the shopping basket created on Jelmoli-shop.ch and the corresponding personalized advertisements on 20min.ch.	40
A.4	Screen shots of the shopping basket created on Laredoute.ch and the corresponding personalized advertisements on Tagesanzeiger.ch.	41
A.5	Screen shots of the shopping basket created on Manor.ch and the corresponding personalized advertisements on 20min.ch.	42
A.6	Screen shots of the shopping basket created on Zalando.ch and the corresponding personalized advertisements on Blick.ch.	43

List of Tables

3.1	Number of involved domains when loading landing page.	16
3.2	Content types when loading landing page.	16
3.3	Number of total second-level domains and second-level ad domains on five news sites.	17
3.4	Percentage of loaded content due to advertisements and percentage of requests to ad domains.	18
4.1	Percentage of same content on same URL for news sites and delivered content type.	21
5.1	Description of browsing profiles created on the 12 online shops.	24
5.2	Shopping sites leading to personalized ads on different news sites.	26

Bibliography

- [1] B. für Statistik Schweiz, “Detaillierte Branchentabellen 2008 - 2012 gemäss NOGA 2008.” <http://www.estv.admin.ch/dokumentation/00075/00076/00714/index.html?lang=de>, October 2014.
- [2] B. für Statistik Schweiz, “Haushalte und Bevölkerung - Online Aktivitäten.” http://www.bfs.admin.ch/bfs/portal/de/index/themen/16/04/key/approche_globale.indicator.30111.301.html, October 2014.
- [3] E. Pariser, *The filter bubble: What the Internet is hiding from you*. Penguin Press HC, 2011.
- [4] “Top Sites in Switzerland.” <http://www.alexa.com/topsites/countries/CH>, October 2014.
- [5] “Shopping-Portale und Onlineshops Schweiz 2014.” <http://blog.carpathia.ch/poster2014/>, October 2014.
- [6] F. Roesner, T. Kohno, and D. Wetherall, “Detecting and Defending Against Third-party Tracking on the Web,” in *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation, NSDI'12*, (Berkeley, CA, USA), pp. 12–12, USENIX Association, 2012.
- [7] A. Schroeter, P. Westermeyer, C. Müller, T. Schlottke, and C. Wendels, “Real-Time Advertising: Funktionsweise, Akteure, Strategien.” <http://rtb-buch.de/>, 2013.
- [8] L. Olejnik, T. Minh-Dung, and C. Castelluccia, “Selling Off Privacy at Auction,” (San Diego, CA, USA), NDSS, November 2014.
- [9] D. Gugelmann, B. Ager, and V. Lenders, “Towards Classifying Third-Party Web Services at Scale,” in *Proceedings of the 2014 CoNEXT on Student Workshop, CoNEXT Student Workshop '14*, (New York, NY, USA), pp. 34–36, ACM, 2014.
- [10] “Mitmdump.” <https://mitmproxy.org/doc/mitmdump.html>, October 2014.
- [11] “Pandas: Python Data Analysis Library.” <http://pandas.pydata.org/>, October 2014.
- [12] “Libmproxy.” <https://mitmproxy.org/doc/scripting/libmproxy.html>, October 2014.
- [13] “SeleniumHQ: Browser Automation.” <http://www.seleniumhq.org/>, October 2014.
- [14] “NetworkX.” <https://networkx.github.io/>, October 2014.