



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Master's Thesis

Vorlage für Abschlussarbeiten

Aram Yesildeniz

max.peter.mustermann@informatik.uni-hamburg.de

Studiengang Informatik

Matr.-Nr. 12345678

Erstgutachter: Professor A. Ersthelfer

Zweitgutachter: Professor Z. Eswirdschonwerden

Abgabe: 01.09.2021

A distributed system is one where the failure of some
computer I've never heard of can keep me from getting my work done.
– *Leslie Lamport*

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 1.1 | E-Commerce | 1 |
| 1.1.1 | User Satisfaction and Performance | 1 |
| 1.2 | Web Analytics | 1 |
| 1.2.1 | Web Performance | 1 |
| 1.2.2 | Tools | 2 |
| 1.3 | Research Question | 2 |
| 2 | Terms and Definitions | 3 |
| 2.1 | Metrics | 3 |
| 2.1.1 | Introduction | 3 |
| 2.1.2 | "Non-Performance" metrics | 3 |
| 2.1.3 | Performance Metrics | 6 |
| 2.1.4 | WebPageTest Metrics | 8 |
| 2.1.5 | Google Analytics Site Speed Metrics | 12 |
| 2.1.6 | Comparison | 13 |
| 2.2 | Measuring Methods | 15 |
| 2.2.1 | Synthetic Monitoring | 15 |
| 2.2.2 | Real-User Monitoring | 15 |
| 3 | Related Work | 17 |
| 3.1 | WebPageTest | 17 |
| 3.1.1 | Overview | 17 |
| 3.1.2 | Configuration | 17 |
| 3.1.3 | Private Instances | 18 |
| 3.2 | Google Analytics | 18 |
| 3.3 | Research | 18 |
| 3.3.1 | some title for first category | 18 |
| 3.3.2 | Research about Tools | 19 |
| 3.3.3 | Research about Metrics | 19 |
| 4 | Approach | 21 |
| 4.1 | Empirical Research Methods | 21 |
| 4.1.1 | Controlled Experiment | 21 |
| 4.1.2 | Test Setup | 21 |

Contents

| | | |
|----------|---|-----------|
| 4.1.3 | Independent Variables within template | 22 |
| 4.2 | Test Object: HTML Template / Test website ideas | 22 |
| 4.2.1 | WordPress | 26 |
| 4.2.2 | Plain / Skeletal Website | 26 |
| 4.2.3 | HTTP Archive inspired website | 26 |
| 4.2.4 | Mirroring a complete e-commerce website | 26 |
| 4.3 | Test Runs | 27 |
| 4.3.1 | WPT Configurations | 27 |
| 4.3.2 | Website Variations | 28 |
| 4.3.3 | Test Plan | 29 |
| 4.3.4 | Tool support for diagrams and data analysis | 30 |
| 5 | Evaluation | 31 |
| 5.1 | Internal, external validity | 31 |
| 5.2 | General | 31 |
| 5.3 | Plain / Skeletal Website | 32 |
| 5.4 | Mirroring | 32 |
| 5.5 | HTTP Archive inspired website | 32 |
| 5.6 | WebPageTest Bulk Tests | 32 |
| 5.6.1 | Bulk Test Overview: Description of test result page | 32 |
| 5.6.2 | Summary File for one Test | 33 |
| 5.6.3 | Aggregate Statistics File | 33 |
| 5.6.4 | Compare Section | 33 |
| 5.7 | Wpt waterfall | 35 |
| 6 | Future Work | 37 |
| 6.1 | Limitations of this thesis | 37 |
| 6.2 | Other measurement tools and metrics | 37 |
| 6.2.1 | Google Analytics 4 | 37 |
| 6.3 | Speed Kit | 37 |
| 6.4 | PWAs, AMPs, Service Workers, Caching, HTTP2 etc. | 37 |
| 7 | Conclusion | 39 |
| 8 | Appendix | 41 |
| 8.1 | WebPageTest Bulk Tests | 41 |
| 8.1.1 | Single Test Raw page data | 41 |
| 8.1.2 | Single Test Raw object data | 45 |
| 8.1.3 | Single Test Http archive (.har) | 45 |
| 8.1.4 | Combined Test Raw page data | 45 |
| 8.1.5 | Combined Test Raw object data | 45 |
| 8.1.6 | Combined Test Aggregate data | 45 |

| | |
|-------------------------------------|-----------|
| Bibliography | 47 |
| Eidesstattliche Versicherung | 49 |

1 Introduction

- Explain structure of this thesis and maybe give some "eckdaten"
- This chapter: Describe shortly all sections from this chapter
- In the next chapter...

1.1 E-Commerce

- Overview and general introduction to e-commerce
- Relevance of e-commerce

1.1.1 User Satisfaction and Performance

- What is user satisfaction and why is it important
- Prove that Page Speed = Money
- Introduction to A/B Testing. How and for which reasons A/B Testing is used in practice.

A/B Testing

1.2 Web Analytics

- Historical background and contextualisation, usage, definition
- Web Analytics Process
- Mechanisms, Measurement methods / Collecting data: Log file analysis, client site page tagging, alternatives
- KPIs ?

1.2.1 Web Performance

- What is web performance? Why it matters
- Overview of factors that impact performance, bottlenecks
- Overview of measurement methods, techniques and metrics

1 Introduction

1.2.2 Tools

- Some short overview about existing tools
- Conclude that I use WPT for synthetic performance testing and GA for RUM

1.3 Research Question

- Difficulty of defining scope
- Measuring performance of a web site impacts its performance or other effects take place / Observer effect
- Why the research question is relevant

2 Terms and Definitions

- Last chapter...
- This chapter: Describe shortly all sections from this chapter
- In the next chapter...
- This chapter should cover all relevant terms and definitions within web performance measurement
- How terms can be structured / taxonomy
- Ambiguity of definitions

2.1 Metrics

2.1.1 Introduction

- Metrics jungle, difficulty of taxonomy
- Performance vs UX

2.1.2 "Non-Performance" metrics

- User engagement: session length, bounce rate, etc.
- Business KPIs: Cart size, conversion rate, etc.
- QA metadata: Page views, JS errors, etc.
- Hit
- Click-Through
- Page View
- Visit
- Visitor / Unique Visitor
- Referrer
- Conversion Rate

2 Terms and Definitions

- Abandonment Rate
- Attrition
- Loyalty, Frequency and Recency
- Measuring Reach: ...
- Measuring Acquisition: ...
- Measuring Conversion: ...
- Measuring Retention: ...
- Basic metrics (see table): basic metrics are meaningless
- Advanced metrics: Customer lifecycle analysis, customer behaviour analysis
- Types: Counts, Ratios, KPIs
- Definitions for all terms, like Page view, unique visitor, etc.
- Importance of setting goals
- Conversion Rate
- Kennzahlen für Websites nach Typ: ROI-Ebene, Online-Shop, ...
- Conversion Rates, pages that visitors abandon most
- Click throughs
- UGC (User generated content)
- Subscriptions, Signups
- Referring URL
- Visitor Motivation, VOC: Voice of the Customer
- Ad and campaign effectiveness
- Findability and Search Effectiveness
- Trouble Ticketing and Escalation
- Loyalty: Ratio of new to returning visitors; average time between visits; time since last login; rate of attrition or disengagement

p.15 "whether your business benefited in some way from their visits."

The percentage of visitors that your site converts to contributors, buyers, or users is the most important metric you can track -> Conversion Rate

- 4 categories: site usage, referrers, site content analysis, quality assurance
- 8 fundamental metrics
- Site usage:
 - Demographics and System Statistics
 - Internal Search Information
 - Visit Length
 - Visitor Type
- Referrers:
 - Referring URL and Keyword Analysis
- Site content analysis:
 - Top Pages
 - Visitor Path
- Quality assurance:
 - Errors
- Erfolg messen und bewerten
- Traffic:
 - Page Impression / Page View
 - Visit
 - Visitor / Unique visitor
- Bounce rate
- Conversion rate
- CTR: Click-through-rate
- Session length
- Good metrics should be: Uncomplex, Relevant, Timely, Instantly Useful
- Basic metrics: Visits, bounce rate, page views, pages/visits, avg time, % new visits
- Guidance Performance Indicator (GPI) metric
- Visit count: page view, visit, unique visitor
- Visit duration: time on page, time on site.

2 Terms and Definitions

- Bounce rate and exit rate.
- Besucheranalyse: Wie viele Besucher?, Anzahl Besucher mit Mobilgerät, Demographische Daten (Geschlecht, Altersgruppe)
- Seitenanalyse: Was machen die Besucher im Shop?, Zielseite / Startseite: Erste Seite, die ein Besucher angeschaut hat, Ausstiegseite
- E-Commerce-Analyse: Transaktions-daten aus Shop, Funnel-Analyse
- Types: Anzahl, Relations, Werte
- Content: Where, Who, How, What
- Hits
- Page Views
- Visits / Sessions
- Visitor / Unique Visitor

2.1.3 Performance Metrics

- Introduction to the Web Performance Working Group
- Overview of Browser APIs and the data they expose: High Resolution Time API, Navigation Timing API, etc.
- If possible make one deep dive into one API: What exactly gets measured? Maybe check out html standard, v8 or chromium implementation, etc.

Standards and APIs, Browser metrics, standards

- Web Performance Working Group
- User Timing API
- Navigation Timing API: Level 1 (performance.timing), Level 2 (PerformanceNavigationTiming) ?
- Network Information API
- Resource Timing API
- Paint Timing API
- High Resolution Time API
- Performance Timeline API

- Performance Observer API
- Long Tasks API
- Element Timing API
- Event Timing API
- Server Timing API

Navigation Timing API

- Show image of navigation timings
- Explain one or two events directly with specification: navigationStart, domInteractive, etc.

Google metrics? User-centric / UX / visual

Web Vitals

- Key questions: Is it usable, is it delightful, ...
- Types of metrics
- important metrics
- custom metrics
- Core Web Vitals: First Input Delay, Cumulative Layout Shift, Largest Contentful Paint
- First Paint, First Contentful Paint: Is it happening? PerformanceObserver
- First Meaningful Paint, Hero Element: Is it useful?
- Time To Interactive: Is it usable? Use Polyfill
- Long Tasks: Is it delightful? PerformanceObserver
- Total Blocking Time
- Time To First Byte

2 Terms and Definitions

Core Web Vitals

- Most important metrics, Apply to all websites, Measures real user experience, Measurement support for Lab and Field, Concise and clear
- LCP: Progressive loading. FCP may become a core web vital
- FID: Interactivity during load
- CLS: Visual stability
- Future goals: Better support for Single Page Apps, Input responsiveness, Scrolling and animations
- Areas of user experience beyond performance: Security, Privacy, Accessibility
- Introduction, what is it
- How to measure
- How to improve
- Introduction, what is it
- How to measure
- How to improve
- Introduction, what is it
- How to measure
- How to improve

Others

- Visually complete ?

Speed Index

2.1.4 WebPageTest Metrics

- Metrics Categories:
 - High Level Metrics:
 - * Document Complete
 - * Fully Loaded
 - * Load Time

- * First Byte
- * Start Render
- * Requests
- * Bytes In (Page Size)
- Page-level Metrics:
 - * Technical Page Metrics:
 - -> APIs, GA Site Speed Metrics
 - TTFB
 - loadTime
 - docTime
 - ...
 - * Visual Metrics:
 - SpeedIndex
 - firstPaint
 - firstContentfulPaint
 - firstMeaningfulPaint
 - ...
 - * Javascript and CPU timings
 - * Page Information
 - * Browser State
 - * Lighthouse Summary Metrics
 - * Optimization Checks/Grades
 - * Instrumented Metrics
 - * Test Information
 - * Misc
- Request-level metrics:
 - * Request Details
 - * Request Timings
 - * Request Stats
 - * Headers
 - * Protocol Information
 - * Javascript/CPU details

2 Terms and Definitions

- * Optimization Checks
- * Misc
- Optimization Grades:
 - Keep-alive Enabled
 - Compress Text
 - Compress Images
 - Cache Static Content
 - Use of CDN
- First View and Repeat View

| Name | Description |
|-------------------|---|
| Successful Tests | Amount of tests who completed successfully |
| Document Complete | The time from the initial request until the browser fires load event. Also known as the document complete time. This is the time at which the Document Object Model (DOM) has been created and all images have been downloaded and displayed. For most traditional web pages, the load time is a suitable metric for representing how long a user must wait until the page becomes usable. This is the default performance metric on WebPageTest. Also known as Load Time (?). Around this time, the page's script is hard at work in the load-event handler firing off more requests for secondary content. The incomplete nature of this metric is why Fully Loaded was added to the table of metrics from the previous section. window.onload (?). The point where the browser onLoad event fires. The equivalent Navigation Timing event is loadEventStart. Document Complete Time: Amount of time that has elapsed from the initial page request until the browser fires the load event. This is the time at which the Document Object Model (DOM) has been created and all images have been downloaded and displayed. |

| | |
|--------------------|--|
| Fully Loaded | The time from the initial request until WebPageTest determines that the page has finished loading content. The page might have waited for the load event to defer loading secondary content. The time it takes to load the secondary content is accounted for in the Fully Loaded Time. The time (in ms) the page took to be fully loaded — e.g., 2 seconds of no network activity after Document Complete. This will usually include any activity that is triggered by javascript after the main page loads. The point after onLoad where network activity has stopped for 2 seconds. Specific to WebPageTest and not provided by Performance API. Fully loaded waits for 2 seconds of no network activity (and no outstanding requests) after onLoad and then calls it done (only measures to the last activity, doesn't include the 2 seconds of silence in the measurement). Fully Loaded is a measure based on the network activity and is the point after onload when there was no activity for 2 seconds. |
| First Byte | Time until the server responds with the first byte of the response. |
| Start Render | Time until the browser paints content to the screen. The time for the browser to display the first pixel of content (paint) on the screen. Time until the browser paints content to the screen. WebPageTest's own metric, determined by programmatically watching for visual changes to the page. Same as First Render? |
| Bytes In (Doc) | Total size of the Document Complete Requests' response bodies in bytes. |
| Requests (Doc) | Number of HTTP requests before the load event, not including the initial request. |
| Load Event Start | Time in ms since navigation started until window.onload event was triggered (from W3C Navigation Timing). |
| Speed Index | See Speed Index |
| Last Visual Change | Time in ms until the last visual change occurred. Last change is a completely visual measurement and is the last point in the test when something visually changed on the screen. It could be something as simple as an animated gif or ad even that didn't really cause much CPU work but changed some pixels on the screen. It is only captured when video is recorded because it depends on the video capture to measure it. |
| Visually Complete | Time in ms when page was visually completed. Is measured from a video capture of the viewport loading and is the point when the visible part of the page first reached 100% "completeness" compared to the end state of the test. |

Table 2.1: Your caption here

2.1.5 Google Analytics Site Speed Metrics

Show with analytics.js that it is indeed those navigation timing api calculations.

Ec = function (a)...

GA does not really provide any UX metrics! The site speed metrics are all from navigation timing api which are measurements from the browser.

GA Site Speed Metrics (description from https://support.google.com/analytics/answer/2383341?hl=en&ref_topic=1282106)

<https://stackoverflow.com/questions/18972615/how-do-the-metrics-of-google-anal>

| Name | Description |
|------------------------------------|--|
| Page Load Sample | The number of pageviews that were sampled to calculate the average page-load time. |
| Speed Metrics Sample | The sample set (or count) of pageviews used to calculate the averages of site speed metrics. This metric is used in all site speed average calculations, including avgDomainLookupTime, avgPageDownloadTime, avgRedirectionTime, avgServerConnectionTime, and avgServerResponseTime. |
| DOM Latency Metrics Sample | Sample set (or count) of pageviews used to calculate the averages for site speed DOM metrics. This metric is used to calculate ga:avgDomContentLoadedTime and ga:avgDomInteractiveTime. |
| Page Load Time (sec) | The average amount of time (in seconds) it takes that page to load, from initiation of the pageview (e.g., click on a page link) to load completion in the browser. |
| Domain Lookup Time (sec) | The average amount of time spent in DNS lookup for the page. |
| Page Download Time (sec) | The time to download your page. |
| Redirection Time (sec) | The time spent in redirection before fetching the page. If there are no redirects, the value for this metric is expected to be 0. |
| Server Connection Time (sec) | The time needed for the user to connect to your server. |
| Server Response Time (sec) | The time for your server to respond to a user request, including the network time from the user's location to your server. |
| Document Interactive Time (sec) | The average time (in seconds) that the browser takes to parse the document (DOMInteractive), including the network time from the user's location to your server. At this time, the user can interact with the Document Object Model even though it is not fully loaded. |
| Document Content Loaded Time (sec) | The average time (in seconds) that the browser takes to parse the document and execute deferred and parser-inserted scripts (DOMContentLoaded), including the network time from the user's location to your server. Parsing of the document is finished, the Document Object Model is ready, but referenced style sheets, images, and sub-frames may not be finished loading. This event is often the starting point for javascript framework execution, e.g., JQuery's onready() callback, etc. |

2.1.6 Comparison

- We can show this with experiments
- Load test page on a specific day only once and save timings exposed by perfor-

| Navigation Timing API | WPT | GA |
|--|-------------------------------------|----------------------|
| loadEventStart - navigationStart | Document Complete, Load Event Start | pageLoadTime |
| domainLookupEnd - domainLookupStart | DNS lookup, dns_ms | domainLookupTime |
| connectEnd - connectStart | connect_ms | serverConnectionTime |
| responseStart - requestStart | .. | serverResponseTime |
| responseEnd - responseStart | .. | pageDownloadTime |
| fetchStart - navigationStart | .. | redirectionTime |
| domInteractive - navigationStart | .. | domInteractiveTime |
| domContentLoadedEventStart - navigationStart | domContentLoadedEventStart | domContentLoadedTime |

mance.timing object (from console)

- Calculate differences corresponding to the table
- Get GA data for that day and save it
-

2.2 Measuring Methods

- Explanation and comparison of synthetic and real-user monitoring with concrete examples
- Short overview of other measuring methods such as log analysis or surveys

2.2.1 Synthetic Monitoring

- What is it
- How does it work
- Application, real life scenario
- Examples:
 - WebPageTest
 - Google Lighthouse
 - Other solutions

2.2.2 Real-User Monitoring

- What is it
- How does it work
- Application, real life scenario
- Examples:
 - Google Analytics
 - CrUX
 - SpeedKit
 - Other solutions

3 Related Work

- Last chapter...
- This chapter: Describe shortly all sections from this chapter
- In the next chapter...
- This chapter should list research which covers and explores questions relevant for this thesis, such as:
 - Metrics: New metrics, meaning of metrics, difficulties of defining metrics, etc.
 - Overview, evaluation and comparison of measurement tools and methods
 - If available: Impact of RUM on performance

3.1 WebPageTest

- Overview
- Configuration
- Private Instances

3.1.1 Overview

- What is it
- Why to use it, Who uses it, how to use it
- Waterfall and Grades

3.1.2 Configuration

- Caching, repeat view
- Traffic shaping

3 Related Work

3.1.3 Private Instances

- Architecture
- AWS
- Docker localhost
- Bulk tests

3.2 Google Analytics

- Custom metrics with Google Web Vitals as example
- Show how to include GA script (analytics.js, gtag, Tag Manager, etc.)
- Show some real life examples how script code is included into page, e.g. from Amazon, Otto etc

3.3 Research

- Research exists about topics like:
- Here i will provide a list of in my eyes relevant papers, summaries them and discuss why this is important for my research

3.3.1 some title for first category

2014 Singal I. - Describes history of web analytics and tools - Provides definitions and taxonomy for metrics - Describes log file vs page tagging - Describes KPIs

II. - Lit. overview for KPIs and Web Metrics - Lit. overview for "Trust" - Lit. overview for "Fuzzy" -> What are does categories?

III. - Some other literature worth mentioning

IV. - Describes 8 open challenges for researchers

2015 Bekavac - Two parts: - 1: Some general overview of web analytics, tools and metrics, KPIs etc - 2: Empirical study about employees satisfaction of used web analytics tools

1: - 9 web business models and 5 common goals - Hypothesis: Web analytics tools track and improve a user's satisfaction with web-based business models. - Web analytics definition. Log files vs Site Tagging - Web Analytics process - Tools: 5 categories, Process of selecting tool, Table with features of different tools - Web metrics categories, Table with business models and their KPIs

2: - Which tools are used for which purpose / Activity - Users satisfaction

3.3.2 Research about Tools

Kaushik 2007 - Provides 3 questions which help to choose web analytics tools

2011 Nakatani - Gives some arguments why web analytics is important for business
- Provides different categorizations for web analytics tools - Gives pros and cons of log file analysis and page tagging - Provides tool selection method based on AHP (Analytic Hierarchy Process)

2016 Kaur

3.3.3 Research about Metrics

- Dont know:

4 Approach

- Last chapter...
- This chapter: Describe shortly all sections from this chapter
- In the next chapter...
- In this chapter the practical work should be documented and explained
- Elaboration of how the practical work could help answer the research question
- Discussion of real-life setup and how experiments approach it

4.1 Empirical Research Methods

- Overview of methods
- reproduceability etc.
- validity
- Justification why following approaches are conducted as controlled experiments

4.1.1 Controlled Experiment

- Short overview about controlled experiments in computer science
- Design: Show test setup image: Independent and dependent variables
- Hypothesis testing

4.1.2 Test Setup

- What is test object (website)
- What are dependent variables: Performance metrics
- What are independent variables: Specific changes in test object (see next chapter)

4 Approach

Measure effects: Dependent Variables

- Performance metrics from Lab and Field, see terms and definitions
- But also quality of RUM data. Because we could have a nice performance but RUM will be of bad quality.

Test object / HTML Template

- Depending on different approaches / Ideas (see next chapter), template looks different
- But general structure stays the same and independent variables can be defined
- Here we show different independent variables and variants

Lab and Field

- I want to collect Lab and field data for dependent variables for comparison
- This setup is a special case because lab bots (e.g. WPT) simulate at the same time real users for RUM data

4.1.3 Independent Variables within template

- IV 1: Position of included analytics script. Values: top-head, bottom-head, bottom-body
- IV 2: Attribute of included analytics script: no-attribute, async, defer
- IV 3: Other tracking script included
- Other IVs not included but worth mentioning

Other IVs not included but worth mentioning

- More or less infinite number of independent variables
- Again the big and important fact that each website is different

4.2 Test Object: HTML Template / Test website ideas

- Several ideas are proposed
- Each idea has pro and contra: each idea should be discussed of its usefulness, advantages and disadvantages

Listing 4.1: Position 1

```
<!DOCTYPE html>
<html>
  <head>
    <!-- Google Analytics -->
    <script></script>
    <!-- End Google Analytics -->

    <title>
    <meta>
    <link>
    <script>
  </head>

  <body>
    ...
  </body>
</html>
```

Listing 4.2: Position 2

```
<!DOCTYPE html>
<html>
  <head>
    <title>
    <meta>
    <link>
    <script>

    <!-- Google Analytics -->
    <script></script>
    <!-- End Google Analytics -->
  </head>

  <body>
    ...
  </body>
</html>
```

Listing 4.3: Position 3

```
<!DOCTYPE html>
<html>
  <head>
    <title>
    <meta>
    <link>
    <script>

  </head>

  <body>
    ...
    <!-- Google Analytics -->
    <script></script>
    <!-- End Google Analytics -->
  </body>
</html>
```

Listing 4.4: Attribute 1

```
<!DOCTYPE html>
<html>
  <head>
    <!-- Google Analytics -->
    <script></script>
    <!-- End Google Analytics -->

    <title>
    <meta>
    <link>
    <script>
  </head>

  <body>
    ...
  </body>
</html>
```

Listing 4.5: Attribute 2

```
<!DOCTYPE html>
<html>
  <head>
    <!-- Google Analytics -->
    <script async></script>
    <!-- End Google Analytics -->

    <title>
    <meta>
    <link>
    <script>
  </head>

  <body>
    ...
  </body>
</html>
```

Listing 4.6: Attribute 3

```
<!DOCTYPE html>
<html>
  <head>
    <!-- Google Analytics -->
    <script defer></script>
    <!-- End Google Analytics -->

    <title>
    <meta>
    <link>
    <script>
  </head>

  <body>
    ...
  </body>
</html>
```

Listing 4.7: Other Script 1

```

<!DOCTYPE html>
<html>
  <head>
    <!-- Google Analytics -->
    <script></script>
    <!-- End Google Analytics -->

    <title>
    <meta>
    <link>
    <script>
  </head>

  <body>
    ...
  </body>
</html>

```

Listing 4.8: Other Script 2

```

<!DOCTYPE html>
<html>
  <head>
    <!-- Google Analytics -->
    <script></script>
    <!-- End Google Analytics -->

    <!-- Other Script -->
    <script></script>
    <!-- End Other Script -->

    <title>
    <meta>
    <link>
    <script>
  </head>

  <body>
    ...
  </body>
</html>

```

4 Approach

4.2.1 WordPress

- Show usage of WordPress with some statistics: Why is it so verbreitet
- Explain Plugin system
- Explain Setup on localhost with wocommerce and GA plugin
- Elaborate why this idea was not used

4.2.2 Plain / Skeletal Website

- Idea: Lab environment to have control over all and see effects of changing independent variables
- Problem: Too far away from reality
- Use this as the simplest test possible, not even POC (POC is http archive site)

4.2.3 HTTP Archive inspired website

- Idea: Get correct page weight
- POC: Show that changing independent variables X affect result

4.2.4 Mirroring a complete e-commerce website

Re-write this to otto start page clone chapter

Manual adjustments: - Move everything to test folder because top domain is /otto

What did not work (mostly 404s): - user-set-consent-id-cookie: Cookie with name consentId is not set, user-set-consent-id-cookie returns therefore 404 - subscribeToNewsletterSnippetContent: Change path did not work... - amount.json: Not found, also wl_miniWishlistAmount in local storage does not created - a_info: Mock a_info response json does not work...

- footer - userTiming

WPT RV is returning empty csv when 404s are encountered. Therefore i mock the missing ressources so that WPT can run bulk tests successfully.

- mock image sprite_all_1ba408b2.png

- create empty file called user-set-consent-id-cookie

- change path for subscribeToNewsletterSnippetContent: This will remove the cookie banner... but then WPT works

- Idea: Close to reality as possible
- Problems when mirroring a website
- Elaborate why this idea of mirroring complete website was not used

- I used mock of start page of otto, which works fine
- Compare original otto website with mock

Comparison to original webpage

- Remove GA again from mock, so that mock and original are as similar as possible
 - Run the same lab test on both pages: WPT and maybe lighthouse
 - Compare both results and explain differences
- Setup: Run WPT on mock and on original website - WPT config: - Browser: Chrome
 - Number of test runs: 3 - FV and RV - Capture Video - Capture DevTools Timeline - Bulk testing: 100x

Diagrams with FV and RV for both cases:

Technical: - First Byte - Bytes In (Doc) - Requests (Doc)

Visual: - Document Complete - Speed Index

Problem with Repeat View - Problem with RV, Caching: Otto sets request headers to cache-control: no-cache which means that RV basically downloads all resources again. The mock is hosted on Github, where the cache-control header is set to ... It is not possible to change the github request headers. We can modify the http request headers via html, but this is not a clean solution. Therefore I use a different e-commerce website which does not shut down caching so that the RV results are more similar.

Ideally I would host the mock website on a similar infrastructure as the original site with the same webserver configuration. This is for a masters thesis not feasible.

4.3 Test Runs

- This section covers all conducted test runs
- Explain test configuration: how many runs, dependent and independent variables, etc.

4.3.1 WPT Configurations

General Settings

Explanations First View: "First View refers to the cold cache setup in which nothing is served locally" Repeat View: "Repeat View refers to the warm cache containing everything instantiated by the first view" (2016 Using WPT p. 62)

Capture Video: ...

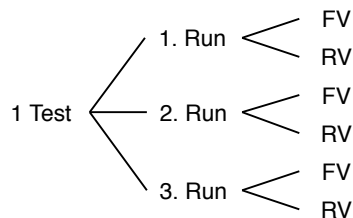
For one test, we have actually six times that the website gets loaded and tested. For e.g. 500 URLs in the bulk test list, we have a total of $500 \times 6 = 3000$ page hits.

4 Approach

Table 4.1: Test Runs [Sch99]

| Configuration Setting | Options | GA |
|--|---|-----|
| Test Location | Test Location | . |
| Browser | Firefox, Chrome | . |
| Connection | LAN | . |
| Number of Tests to Run | 1 to 9 | .. |
| Repeat View | First View and Repeat View, First View Only | . |
| Capture Video | True or False | .. |
| Keep Test Private | True or False | ... |
| Label | Any String | ... |
| Advanced Tab | ... | ... |
| Chromium Tab | ... | ... |
| Auth, Script, Block, SPOF, Custom Tabs | ... | ... |
| Bulk Testing Tab | List of URLs | ... |

Figure 4.1: Number of tests to run: 3, First View and Repeat View



Configuration 1

Configuration 2 Emulate Mobile Browser

Traffic Shaping

- Important to have stable and realistic network condition
- Chromes tool is not the best for this
- Private WPT Instance docker on mac does not allow traffic shaping functionality from WPT
- I use Network Link Conditioner from Apple to slow down the whole machine. See in same blogpost that Patrick highly recommends this
- WPT also slows down their whole machines
- IN general internet connection is very unstable. If i run network link conditioner with e.g. DSL each speedtest gives different results. And other test platforms such as fast.com gives also different result.

Table 4.2: Configuration 1

| Configuration Setting | Option |
|--|---------------------------------|
| Test Location | Test Location |
| Browser | Chrome |
| Connection | LAN |
| Number of Tests to Run | 3 |
| Repeat View | First View and Repeat View |
| Capture Video | True |
| Keep Test Private | False |
| Label | none |
| Advanced Tab | Nothing selected |
| Chromium Tab | Nothing selected |
| Auth, Script, Block, SPOF, Custom Tabs | Nothing |
| Bulk Testing Tab | URL aramyесildeniz... 100 times |

4.3.2 Website Variations

Positions: 1: Top of head element 2: just before closing head element 3: just before closing body element

| Variant | Attribute | Position | Additional Scripts |
|-----------|-----------|----------|--------------------|
| Variant 1 | none | 1 | no |
| Variant 3 | async | 1 | no |
| | defer | 1 | no |
| | none | 2 | no |
| | async | 2 | no |
| | defer | 2 | no |
| | none | 3 | no |
| | async | 3 | no |
| | defer | 3 | no |
| Variant 2 | none | 1 | yes |
| | async | 1 | yes |
| | defer | 1 | yes |
| | none | 2 | yes |
| | async | 2 | yes |
| | defer | 2 | yes |
| | none | 3 | yes |
| | async | 3 | yes |
| | defer | 3 | yes |

Table 4.3: Your caption here

variant1.html

variant2.html

variant3.html

4 Approach

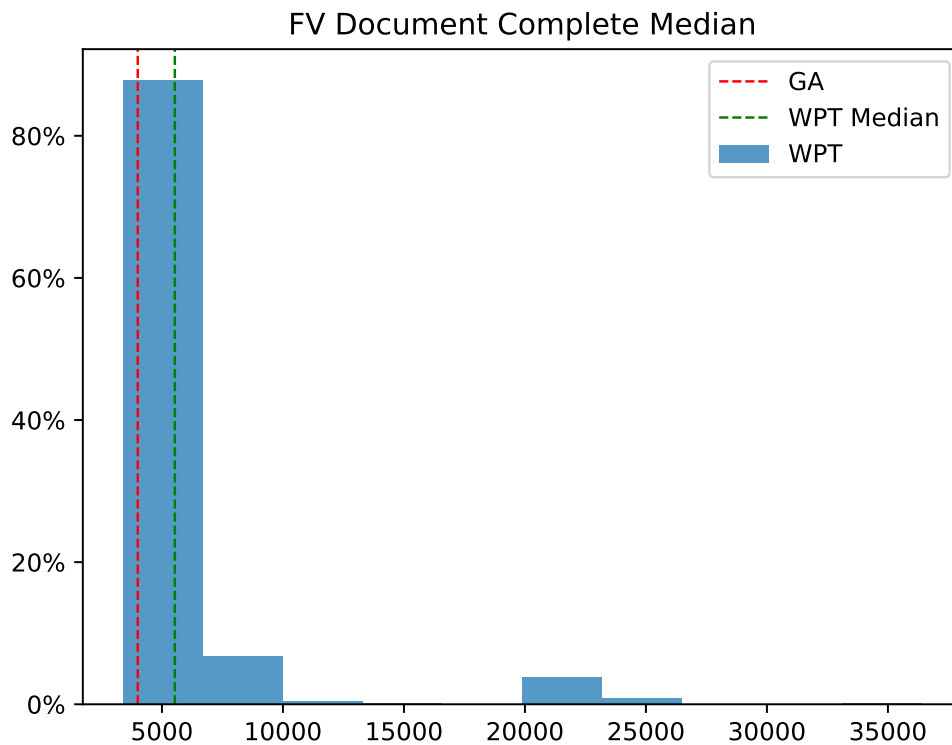
Table 4.4: Test Runs [Sch99]

| WPT | Website | Date |
|-----|---------|------------|
| C1 | V1 | 2021-04-01 |
| .. | .. | .. |

4.3.4 Tool support for diagrams and data analysis

- python
- Matplotlib

5 Evaluation



- Last chapter...
- This chapter: Describe shortly all sections from this chapter
- In the next chapter...

5.1 Internal, external validity

5.2 General

- For each attempt, describe: Threats to validity, generalizability

generalizability: meine Daten zeige nur für Chrome, MacBook, diese Geschwindigkeit etc. Und auch nur für diese Test-Website Die Schwierigkeit der Generalisierbarkeit ist eines der grössten Probleme bei dieser Fragestellung

5.3 Plain / Skeletal Website

- Information gained from this experiment
- Limitations and questions which can not be answered with this approach

5.4 Mirroring

5.5 HTTP Archive inspired website

- Information gained from this experiment
- Meaning and interpretation of the collected data
- Limitations and questions which can not be answered with this approach

5.6 WebPageTest Bulk Tests

- Bulk testing is a feature for private instances only
- Misuse this feature to test the same website X times

5.6.1 Bulk Test Overview: Description of test result page

- Each test has Test ID: YYMMDD_random_random
- Test results after bulk test available under `http://localhost:4000/result/{testID}/`
- For each test run, following data is available:
 - Link to test results: Test result page as same as for single test run
 - Median load time (First view)
 - Median load time (Repeat view)
 - Median Speed Index (First View)
 - Raw page data (file: [TestID_summary.csv])
 - Raw object data (file: [TestID_details.csv])
 - Http archive (.har) (file: json)
- Average First View Load Time
- Average Repeat View Load Time
- Combined Raw: Page Data (file: [TestID_summary.csv])

- Combined Raw: Object Data (file: [TestID_details.csv]). For 100 test runs, this file is appr. 20 MB, 24432 rows, 76 columns.
- Aggregate Statistics (file: [TestID_aggregate.csv])

5.6.2 Summary File for one Test

- Contains 6 rows: 3 test runs: for each test runs 1x first view and 1x repeat view
- Rows 1, 3, 5 contain FV, rows 2, 4, 6 contain data for RV

5.6.3 Aggregate Statistics File

- Contains aggregated data from bulk test
- One row for each test run: For 100 URLs in bulk test will be 100 rows in csv
- Each metric is available with Median, Average, Standard Deviation, Min, Max
- Metrics are available once from FV and once for Repeat View
- Metrics:
 - Successful Tests
 - Document Complete
 - Fully Loaded
 - First Byte
 - Start Render
 - Bytes In (Doc)
 - Requests (Doc)
 - Load Event Start
 - Speed Index
 - Last Visual Change
 - Visually Complete
- => For metric details, see Terms and Definitions

5.6.4 Compare Section

WPT has a feature to compare multiple tests. Accessible under compare URL: <http://localhost:4000/video/compare.php?tests={TestID},{TestID},...>

The compare page contains:

- Film strip

5 Evaluation

- Waterfall diagram
- Visual Progress diagram
- Timings diagram:
 - Visually Complete (First View Visually Complete Median)
 - Last Visual Change
 - Load Time (onload)
 - ...
- Cumulative Layout Shift diagram
- Requests diagram
- Bytes diagram
- Visually complete
- Last Visual Change
- Load Time (onload)
- Load Time (Fully Loaded)
- DOM Content Loaded
- Speed Index
- Time to First Byte
- Time to Title
- Time to Start Render
- CPU Busy Time
- 85% Visually Complete
- 90% Visually Complete
- 95% Visually Complete
- 99% Visually Complete
- First Contentful Paint
- First Meaningful Paint
- Largest Contentful Paint
- Cumulative Layout Shift

- html Requests
- html Bytes
- js Requests
- js Bytes
- css Requests
- css Bytes
- image Requests
- image Bytes
- flash Requests
- flash Bytes
- font Requests
- font Bytes
- video Requests
- video Bytes
- other Requests
- other Bytes

5.7 Wpt waterfall

6 Future Work

- Last chapter...
- This chapter: Describe shortly all sections from this chapter
- In the next chapter...

6.1 Limitations of this thesis

- Discussion of unobserved topics
- Discussion of possible next steps

6.2 Other measurement tools and metrics

- List of tools and metrics worth investigating

6.2.1 Google Analytics 4

6.3 Speed Kit

6.4 PWAs, AMPs, Service Workers, Caching, HTTP2 etc.

- Overview of other web technologies and how they could be relevant for further research

7 Conclusion

- Last chapter...
 - This chapter: Describe shortly all sections from this chapter
 - Scope and contribution of this thesis
 - Short summary of each chapter:
 - Problem statement and why it is worth to examine research question
 - Terms and definitions
 - (Related work)
 - Approach and evaluation of practical work
 - Future work
- Several topics wurden bearbeitet in this thesis, such like mocking a website for testing purposes, literature review, metrics taxonomy, and the main part which is an experiment

8 Appendix

8.1 WebPageTest Bulk Tests

8.1.1 Single Test Raw page data

WPT Metrics from summary file

| Name | Description |
|----------------------------|---|
| minify_total | Total bytes of minifiable text static assets. |
| responses_200 | The number of responses with HTTP status code of 200, OK. |
| testStartOffset | ... |
| bytesOut | The total bytes sent from the browser to other servers. |
| gzip_savings | Total bytes of compressed responses. |
| requestsFull | ... |
| start_epoch | ... |
| connections | The number of connections used. |
| base_page_cdn | The CDN provider for the base page. |
| bytesOutDoc | Same as bytesOut but only includes bytes until the Document Complete event. Usually when all the page content has loaded (window.onload). |
| result | Test result code. |
| final_base_page_request_id | ... |
| basePageSSLTime | ... |
| docTime | Same as loadTime. |
| domContentLoadedEventEnd | Time in ms since navigation started until document DOMContentLoaded event finished. |
| image_savings | Total bytes of compressed images. |
| requestsDoc | The number of requests until Document Complete event. |
| firstMeaningfulPaint | ... |
| score_cookies | WebPageTest performance review score for not using cookies on static assets. |
| firstPaint | RUM First Paint Time, the time in ms when browser first painted something on screen. It's calculated on the client for browsers that implement this method. |
| score_cdn | WebPageTest performance review score for using CDN for all static assets. |
| optimization_checked | Whether or not optimizations were checked. |

8 Appendix

| | |
|----------------------------|---|
| score_minify | WebPageTest performance review score for minifying text static assets. |
| gzip_total | Total bytes of compressible responses. |
| responses_404 | The number of responses with HTTP status code of 404, not found. |
| loadTime | The total time taken to load the page (window.onload) in ms. |
| URL | The tested page URL. |
| score_combine | WebPageTest performance review score for bundling JavaScript and/or CSS assets. |
| firstContentfulPaint | ... |
| image_total | Total bytes of images. |
| score_etags | WebPageTest performance review score for disabling *ETag*s. |
| loadEventStart | Time in ms since navigation started until window.onload event was triggered (from W3C Navigation Timing). |
| minify_savings | Total bytes of minified text static assets. |
| score_progressive_jpeg | WebPageTest performance review score for using progressive JPEG. |
| domInteractive | ... |
| score_gzip | WebPageTest performance review score for using gzip compression for transferring compressable responses. |
| score_compress | WebPageTest performance review score for compressing images. |
| domContentLoadedEventStart | Time in ms since navigation started until document DOMContentLoaded event was triggered (from W3C Navigation Timing). |
| final_url | ... |
| bytesInDoc | Same as byteIn but only includes bytes until Document Complete event. |
| firstImagePaint | ... |
| score_keep-alive | WebPageTest performance review score for using persistent connections. |
| loadEventEnd | Time in ms since navigation started until window.onload event finished. |
| cached | 0 for first view or 1 for repeat view. |
| score_cache | WebPageTest performance review score for leveraging browser caching of static assets. |
| responses_other | The number of responses with HTTPS status code different from 200 or 404. |
| main_frame | ... |
| fullyLoaded | The time (in ms) the page took to be fully loaded — e.g., 2 seconds of no network activity after Document Complete. This will usually include any activity that is triggered by javascript after the main page loads. |
| requests | List of details of all requests on tested page. |

| | |
|---|---|
| final_base_page_request | ... |
| TTFB | Time to first byte, which is the duration in ms from when the user first made the HTTP request to the very first byte of the page being received by the browser. |
| bytesIn | The amount of data that browser had to download in order to load the page. It is also commonly referred to as the page size. |
| osPlatform | ... |
| test_run_time_ms | ... |
| tester | The ID of tester that performed the page test. |
| browser_version | The browser version. |
| document_origin | ... |
| document_URL | ... |
| date | Time and date (number of seconds since Epoch) when test was complete. |
| PerformancePaintTiming.first-paint | ... |
| osVersion | ... |
| domElements | The total number of DOM elements. |
| browserVersion | The browser version. |
| fullyLoadedCPUms | CPU busy time in ms until page was fully loaded. |
| browser_name | The browser name. |
| PerformancePaintTiming.first-contentful-paint | ... |
| base_page_cname | ... |
| eventName | ... |
| os_version | ... |
| base_page_dns_server | ... |
| fullyLoadedCPUpct | Average CPU utilization up until page is fully loaded. |
| domComplete | ... |
| base_page_ip_ptr | ... |
| document_hostname | ... |
| lastVisualChange | Time in ms until the last visual changed occurred. |
| visualComplete | Time in ms when page was visually completed. |
| render | The first point in time (in ms) that something was displayed to the screen. Before that user was staring at a blank page. This does not necessarily mean the user saw the page content — it could just be something as simple as a background color — but it is the first indication of something happening for the user. |
| SpeedIndex | The SpeedIndex score. |
| visualComplete85 | Time in ms when page was visually completed 85%. |
| visualComplete90 | Time in ms when page was visually completed 90%. |
| visualComplete95 | Time in ms when page was visually completed 95%. |
| visualComplete99 | Time in ms when page was visually completed 99%. |
| LargestContentfulPaintType | ... |
| LargestContentfulPaintNodeType | ... |

8 Appendix

| | |
|--|--|
| chromeUserTiming.navigationStart | ... |
| chromeUserTiming.fetchStart | ... |
| chromeUserTiming.responseEnd | ... |
| chromeUserTiming.domLoading | ... |
| chromeUserTiming.markAsMainFrame | ... |
| chromeUserTiming.domInteractive | ... |
| chromeUserTiming.domContentLoadedEventStart | ... |
| chromeUserTiming.domContentLoadedEventEnd | ... |
| chromeUserTiming.firstPaint | ... |
| chromeUserTiming.firstContentfulPaint | ... |
| chromeUserTiming.firstImagePaint | ... |
| chromeUserTiming.firstMeaningfulPaint | ... |
| chromeUserTiming.firstMeaningfulPaintCandidate | ... |
| chromeUserTiming.domComplete | ... |
| chromeUserTiming.loadEventStart | ... |
| chromeUserTiming.loadEventEnd | ... |
| chromeUserTiming.LargestContentfulPaint | ... |
| chromeUserTiming.LargestTextPaint | ... |
| chromeUserTiming.CumulativeLayoutShift | ... |
| run | The run number. |
| step | ... |
| effectiveBps | Bytes per seconds, i.e.: total of bytes in / total time to load the page. |
| effectiveBpsDoc | Same as effectiveBps but until Document Complete event. |
| domTime | The total time in ms until a given DOM element (specified via domelement parameter when running a test) was found on the page. |
| aft | Above the Fold Time (no longer supported). The time taken to load everything in the viewport above the fold. |
| titleTime | Total time in ms until page title was set on browser. |
| domLoading | ... |
| server_rtt | ... |
| smallImageCount | ... |
| bigImageCount | ... |
| maybeCaptcha | ... |
| bytes.html | ... |
| requests.html | ... |
| bytesUncompressed.html | ... |
| bytes.js | ... |
| requests.js | ... |
| bytesUncompressed.js | ... |
| bytes.css | ... |
| requests.css | ... |
| bytesUncompressed.css | ... |
| bytes.image | ... |
| requests.image | ... |

| | |
|----------------------------------|-----|
| bytesUncompressed.image | ... |
| bytes.flash | ... |
| requests.flash | ... |
| bytesUncompressed.flash | ... |
| bytes.font | ... |
| requests.font | ... |
| bytesUncompressed.font | ... |
| bytes.video | ... |
| requests.video | ... |
| bytesUncompressed.video | ... |
| bytes.other | ... |
| requests.other | ... |
| bytesUncompressed.other | ... |
| id | ... |
| chromeUserTiming.InteractiveTime | ... |

Table 8.1: Your caption here

- 8.1.2 Single Test Raw object data
- 8.1.3 Single Test Http archive (.har)
- 8.1.4 Combined Test Raw page data
- 8.1.5 Combined Test Raw object data
- 8.1.6 Combined Test Aggregate data

Bibliography

- [Cer02] CERAMI, Ethan ; ST.LAURENT, Simon (Hrsg.): *Web Services Essentials*. O'Reilly, 2002
- [Gra81] GRAY, Jim: The Transaction Concept: Virtues and Limitations. In: *Very Large Data Bases, 7th International Conference, September 9-11, Cannes, France, Proceedings*, IEEE Computer Society, 1981, 144-154
- [Sch99] SCHWARZ, Kerstin: *DISDBIS*. Bd. 64: *Das Konzept der Transaktionshülle zur konsistenten Spezifikation von Abhängigkeiten in komplexen Anwendungen*. Infix Verlag, St. Augustin, Germany, 1999
- [Wik16] WIKIPEDIA: *Wissenschaftliche Arbeit*. https://de.wikipedia.org/w/index.php?title=Wissenschaftliche_Arbeit&oldid=156007167.
Version: 2016, Abruf: 2016-07-21

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe angefertigt und mich anderer als der im beigefügten Verzeichnis angegebenen Hilfsmittel nicht bedient habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Arbeit vorher nicht in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

Ich bin mit einer Einstellung in den Bestand der Bibliothek des Fachbereiches einverstanden.

Hamburg, den _____ Unterschrift: _____