Usability Testing the d3-hypertree Hierarchy Browser

Christopher Oser

Usability Testing the d3-hypertree Hierarchy Browser

Christopher Oser

Bachelor's Thesis

to achieve the university degree of

Bachelor of Science

Bachelor's Degree Programme: Computer Science

submitted to

Graz University of Technology

Supervisor

Ao.Univ.-Prof. Dr. Keith Andrews
Institute of Interactive Systems and Data Science (ISDS)

Graz, 17 Jun 2019

Benutzerfreundlichkeitstestung des d3-hypertree Hierarchy Browser

Christopher Oser

Bachelorarbeit

für den akademischen Grad

Bachelor der Wissenschaft

Bachelorstudium: Informatik

an der

Technischen Universität Graz

Begutachter

Ao.Univ.-Prof. Dr. Keith Andrews
Institute of Interactive Systems and Data Science (ISDS)

Graz, 17 Jun 2019

Diese Arbeit ist in englischer Sprache verfasst.

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The document uploaded to TUGRAZonline is identical to the present thesis.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Dokument ist mit der vorliegenden Arbeit identisch.

Date/Datum

Signature/Unterschrift

Abstract

This thesis documents the process of testing a web application with the goal of finding relevant usability issues. The application that was tested was authored by Michael Glatzhofer [2019b] and is part of the Master's thesis "Hyperbolic Browsing: Scalable Hierarchy Browsing in Hyperbolic Space".

Testing was done using the Thinking Aloud test method, a popular method for testing software in terms of usability. This thesis starts by describing some origins of the method, followed by a brief explanation of the d3-hypertree and hyperbolic browsing. Then, the exact procedure and how it was adapted and applied for this specific project will be described.

In the second half of the thesis the results of the tests will be summarized. These consist of the user's thoughts and remarks as well as what was observed during the tests. Lastly, some possible improvements gathered from insight gained during the project will be proposed.

Kurzfassung

Diese Arbeit dokumentiert den Ablauf des Testens einer web Applikation in Anbetracht auf Benutzerfreundlichkeit. Die Applikation, die getestet wurde, stammt von Michael Glatzhofer [2019b] und war ein Teil der Masterarbeit „Hyperbolic Browsing: Scalable Hierarchy Browsing in Hyperbolic Space“.

Die Testungen wurden mithilfe der „Thinking Aloud“ Testmethode durchgeführt, eine populäre Methode um Software auf Benutzerfreundlichkeit zu prüfen. Diese Arbeit befasst sich zunächst mit den Ursprüngen der Methode und erklärt kurz „Hyperbolic Browsing“ das im d3-hypertree verwendet wird. Daraufhin wird der genaue Ablauf der Test-Methode beschrieben sowie auch wie er angepasst wurde für diese spezifische Aufgabe.

In der zweiten Hälfte der Arbeit werden die eigentlichen Ergebnisse der Testungen zusammengefasst. Diese beruhen auf den Meinungen und Anmerkungen der Benutzer sowie auf dem was während der Testungen beobachtet wurde. Zuletzt werden mögliche Verbesserungen vorgeschlagen die über die gesamte Arbeit hin gesammelt wurden.

Contents

List of Figures

List of Tables

Acknowledgements

I want to thank my advisor, Keith Andrews, for providing help whenever it was needed as well as enabling a swift execution of this thesis.

Furthermore I want to thank Michael Glatzhofer for providing support and insight into his master's thesis which was very helpful when analyzing it.

I also want to thank my mother, a strong woman without whom I wouldn't be here.

Many thanks also go out to my grand-parents who have supported me up to this day and most likely future days.

Lastly I want to thank music for always supporting me throughout all my endeavors. Special mentions for this thesis go out to Eminem, Saint Jhn and Boogie for keeping my mood balanced while working on this paper no matter the situation.

Christopher Oser
Graz, Austria, 17 Jun 2019

Credits

I would like to thank the following individuals and organisations for permission to use their material:

- The thesis was written using Keith Andrews' skeleton thesis [Andrews 2019].
- Figure 3.1 is used with kind permission of Ramana Rao, Xerox PARC.
- Figure 3.2 is a screenshot taken by and used with kind permission from Keith Andrews.
- Figures 4.3, 4.4, 4.5 and 4.6 were all based on templates provided by Keith Andrews.

Chapter 1

Introduction

This thesis describes a thinking aloud test of a user interface for browsing large hierarchies, known as a hyperbolic tree browser. The thesis first summarizes the thinking aloud test method as well as the process of testing a specific web application in terms of usability. The results of the study are presented towards the end of the thesis.

Chapter 2 gives context to the testing process that was used. The basic procedure as well as the historical background are explained. In Chapter 3 the concept of hyperbolic browsing, which is used in the tested application, is described. Some more background information and history on the subject is summarized to achieve a good knowledge basis before delving into the actual testing.

The second half of the thesis describes the actual testing that was done. Chapter 4 gives insight to the actual testing procedure that was executed during this project and also sums up the testing environment and materials. The final Chapter 5 portrays the findings and results of the tests. The results are grouped by severity and type to obtain a solid overview and come to a conclusion.

Chapter 2

Thinking Aloud Testing

A thinking aloud test is a way of testing the usability of computer applications, websites, or any type of user interface. Each test user is asked to perform certain tasks while at the same time expressing their thoughts verbally [Andrews 2018]. The screen and test user are recorded with the help of software and a video camera, so that the footage can be reviewed and analyzed later on. Through this, it is possible to closely analyze the behavior and the decisions of the test user. The method is very good at finding usability problems, as well as giving concrete examples of what led to these problems in the form of video footage. It requires very little materials or resources and is therefore cheap and simple to set up.

Since the issues or problems encountered by any one test user may be subjective, a certain number of users should be tested in order to be relevant. The tests for this thesis were done with 10 test users, of which one test was a pilot test and the other nine represented the real test. As Jakob Nielsen described in 1994, four to five test users suffice to produce good results and usable findings [Nielsen 1994a]. The benefits of performing a thinking aloud test strongly outweigh the costs in pretty much any scenario.

Thinking aloud testing has some potential issues. The user is slowed down by a little over 17% when thinking aloud and thinking aloud might sometimes influence a user's problem-solving behavior. Both of these considerations mean that it makes no sense to take timing measurements during a thinking aloud test. The thinking aloud testing method is described in further detail by Carol Barnum [2010].

2.1 History

The thinking aloud method has its roots in psychology and was later adopted by computer scientists and tech companies. One of the first mentions of a thinking aloud testing method in regards to software was by Clayton C. H. Lewis [1982] while working for IBM. This was later further explained by C. Lewis and Rieman [1993] in an article. The method was based on work by Ericsson and Simon in psychology. They published multiple papers about the thinking aloud method [Ericsson and Simon 1980; Ericsson and Simon 1984; Ericsson and Simon 1987]. The methods presented by Ericsson and Simon had to be adapted to fit the needs of computer scientists, but proved to be very effective.

Another noteworthy pioneer in terms of usability testing is Jakob Nielsen who published multiple papers on the thinking aloud method [Nielsen 1994a; Nielsen 1994b]. Since then, the method has changed and been adapted by many different people in industry and nowadays the method is regarded as a standard testing method for software usability [Van Den Haak et al. 2003; Krahmer and Ummelen 2004; Cooke 2010]. A reflection and summary on thinking aloud tests and their history can be found in in Janni Nielsen et al. [2002].

Another interesting aspect of the thinking aloud method is its adaptation to blind or sight impaired users of screen readers. This seems to be rather successful as the results were comparable to test users with full sight [Borsci and Federici 2009; Federici et al. 2010].

Chapter 3

Hyperbolic Browsing

Hyperbolic browsing is a data visualization technique for navigation and exploration of large hierarchies. Large hierarchies are difficult to display with traditional methods, since they quickly become overwhelming. In a hyperbolic browser, the tree strucutre is laid out in hyperbolic space, then projected to a 2D or 3D display space. In theory, the entire tree is always displayed, but in practice, nodes around the edges of the current display become so small as to be no longer visible. Nodes and subtrees of interest are explored by dragging them towards the central region of the display, where they are allocated more space. With a simple flick of the wrist, entire subtrees can be fanned out and explored interactively.

Hyperbolic geometry is defined by modifying the fifth of the five euclidean axioms [Forder 1958]. These type of modifications were first introduced in the 19th century by János Bolyai and Nikolai Lobachevsky [Stillwell 1996]. Michael Glatzhofer [2019b] gives a very in-depth explanation about the mathematics behind hyperbolic browsing.

3.1 History

The first hyperbolic browser was presented by Lamping and Rao [Lamping and Rao 1994; Lamping et al. 1995; Lamping and Rao 1996]. This browser was implemented in multiple versions using Common Lisp, C++, and also Java. A screenshot of the original browser can be seen in Figure 3.1.

Another implementation of a hyperbolic browser, H3, was developed by Tamara Munzner [Munzner 1998b; Munzner 1998a]. This time an entire graph viewing application was constructed which used hyperbolic space to display the data. Furthermore, three-dimensional hyperbolic space was used here rather than the two-dimensional space used by Lamping and Rao. A screenshot from this browser can be seen in Figure 3.2.

After this, a multitude of new implementations were introduced. They differed by implementation language and/or functionality. Michael Glatzhofer [2019b] gives an extensive summary of previous implementations of hyperbolic browsers.

3.2 The d3-hypertree

The implementation of hyperbolic browser tested in this thesis is called d3-hypertree and was implemented by Michael Glatzhofer [2019b]. It is a web-based application written in JavaScript, that can easily be embedded into web pages [Glatzhofer 2019a]. One application of d3-hypertree has been the Open Tree of Life [Hinchliff et al. 2015; Rees and Cranston 2017], an ongoing project authored by biologists. It groups all organisms, living or extinct, into a tree of species. The initial view of the Hyperbolic Tree of Life Mammal subtree was used for the tests in this thesis and can be viewed in Figure 3.3. Since

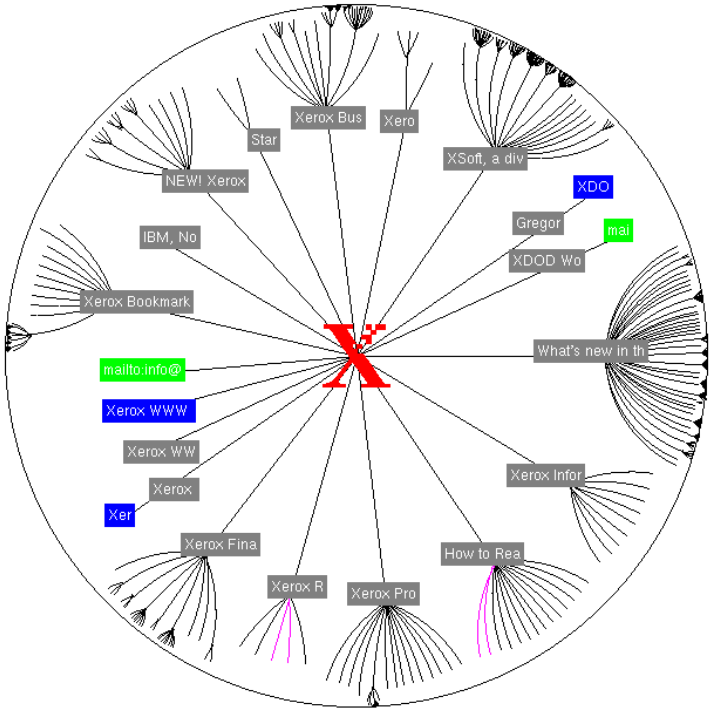


Figure 3.1: The oroginal hyperbolic browser implementation by Lamping & Rao. [Image used with kind permission of Ramana Rao, Xerox PARC.]

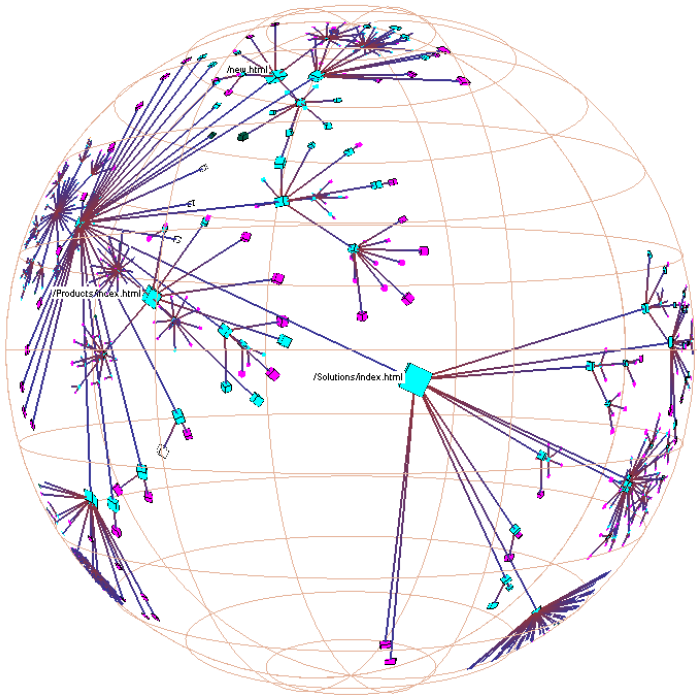


Figure 3.2: The H3 hyperbolic browser by Tamara Munzner. [Image used with kind permission of Keith Andrews.]

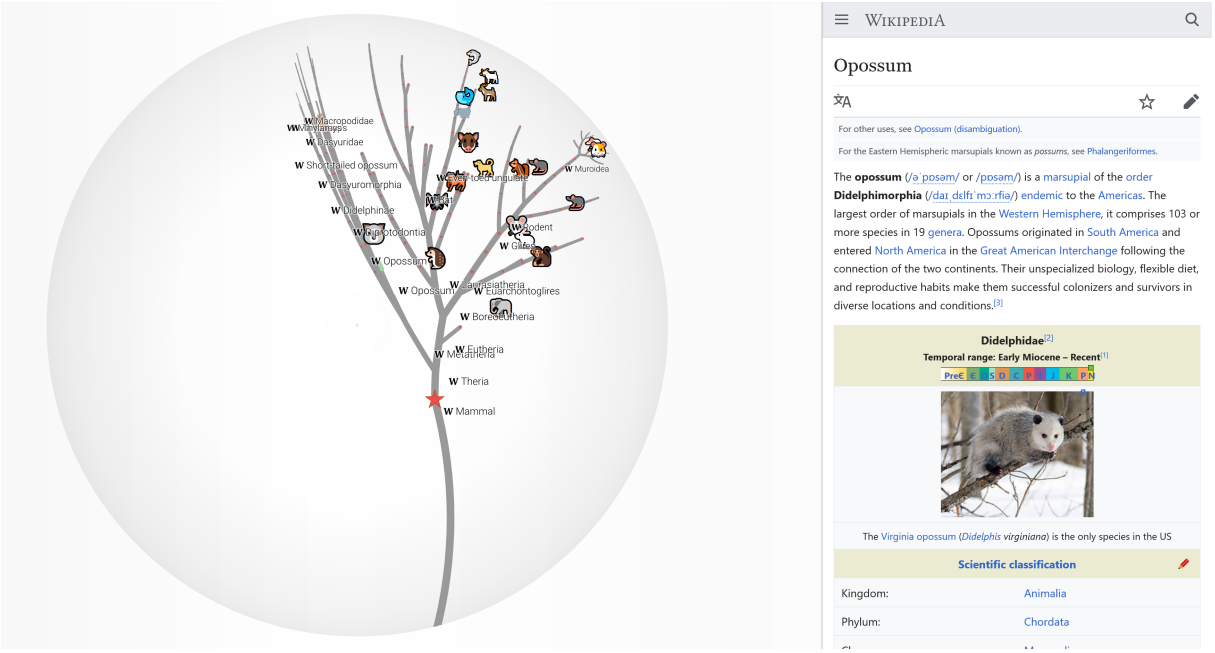


Figure 3.3: The initial view of the Mammal subtree of the Open Tree of Life in d3-hypertree, from the test implementation [Glatzhofer 2018a]. Nodes in the hyperbolic tree are linked to corresponding pages in Wikipedia. [Image used with kind permission of Keith Andrews.]

this is not the only dataset the application can be used on, a different configuration of the d3-hypertree using Google's product taxonomy can be seen in Figure 3.4 [Google 2015]. This demonstrates how the d3-hypertree component can be configured according to one's wishes and needs in order for the data to be displayed optimally.

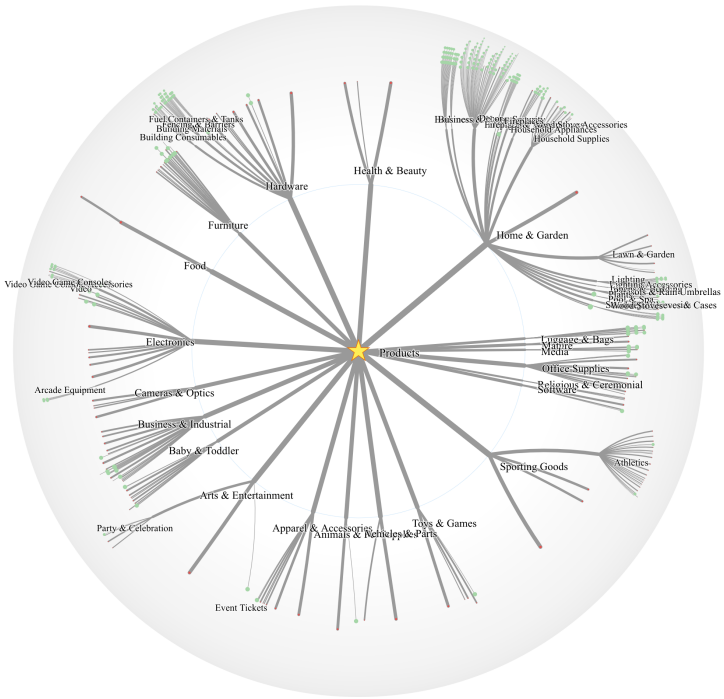


Figure 3.4: The initial view of Google's product taxonomy [Google 2015] in d3-hypertree, from the test implementation [Oser and Glatzhofer 2018]. [Image used with kind permission of Keith Andrews.]

Chapter 4

Test Procedure

The d3-hypertree was tested using the thinking aloud test method in order to find issues or problems concerning the usability of the application. Two different datasets were tested, with slightly different configurations of d3-hypertree: the Mammal subtree of the Hyperbolic Tree of Life [Glatzhofer 2018b] and Google's product taxonomy [Google 2015].

In addition the application was tested on two different devices (tablet and desktop), so differences between platforms would become apparent.

4.1 Datasets

The first dataset that was used was provided by Michael Glatzhofer's main application of the d3-hypertree, the Hyperbolic Tree of Life [Glatzhofer 2018b], based on the Open Tree of Life [Hinchliff et al. 2015; Rees and Cranston 2017]. The Hyperbolic Tree of Life provides a hyperbolic tree visualisation of the Open Tree of Life tree of species and integrates Wikipedia pages of corresponding nodes. The d3-hypertree is set up to use a bottom-up layout, with the root node at the bottom of the panel and the tree growing upwards and outwards like a real-life tree. For this thesis, only the Mammal subtree was used and was hosted separately, in order to provide a smaller dataset for the test users who had no special biological knowledge around subject matter prior to the test. The Mammal subtree contains 13,623 nodes with a maximum depth of 42 levels. Nevertheless, the integration of Wikipedia and the general visual representation were kept the same as in the original Hyperbolic Tree of Life implementation. This smaller dataset will hereafter be referred to as the Tree of Life.

The second dataset was constructed for this thesis using data from the Google Product Taxonomy [Google 2015]. It consists of 5,427 nodes with a maximum depth of 7 levels. The configuration of the product tree is different from the Hyperbolic Tree of Life implementation, in order to see how different visual representations influence the user. The tree spreads over 360 degrees opposed to a much narrower angle in the Hyperbolic Tree of Life. The node labels stick to the nodes instead of floating and only being connected by an indicator line [Oser and Glatzhofer 2018]. This dataset will hereafter be referred to as the Product Tree.

4.2 Users

The spectrum of users who could potentially use the d3-hypertree is very diverse since it heavily depends on the dataset used. For instance, Michael Glatzhofer's Hyperbolic Tree of Life will most likely be used by biology focused people, since most of the tree consists of the Latin names of different species [Glatzhofer 2018b].

Figure 4.1: The desktop PC test setup [Photo taken by the author of this thesis, Christopher Oser].

The tests for this thesis were done with users in the university student age range. This age range was chosen for two reasons. Firstly this is a demographic which is very relevant for the initial users of the d3-hypertree, since the university environment is where it comes from. Secondly, this was the easiest group of test users to acquire at the time. While all of the test users are studying or have studied at universities, they come from different fields including law, marketing, architecture and physiotherapy. This made for a good mix of test users in terms of prior knowledge.

All in all, 10 users completed all the tests, one test user was used in a pilot test, in order to try out the procedure before the remaining users were included. Table 4.1 summarizes the background information gathered from the test users.

4.3 Test Environment

The same test environment was used for each test user. The room was set up while keeping in mind to remove external influences or distractions. For both tests, the user sat down at a simple desk with the device set up accordingly. Additionally a task sheet containing the 5 tasks that had to be fulfilled were placed next to the device.

Two devices were used for the tests: a desktop computer running Windows 10 Pro and an Apple iPad Air 2 running iOS 12.1.1. The Google Chrome web browser version 71 was used on both deviices. The setups can be seen in Figure 4.1 and Figure 4.2. A fresh install of the browser was used and no ad-blockers or other add-ons were used.

4.4 Test Recording

Each session produced two videos. One is a screen recording of the respective device including an audio recording of the test users voice. The other is a camera recording set up behind the user on a tripod in order to capture the test user's behavior during the tests. In addition, mirrors were set on the desk next to the user to capture the test user's facial expressions while completing the tasks.

4.5 Training

No special training on the interface was given to the users, in order to test the intuitiveness of the application for users with no prior knowledge of it. The thinking aloud method was explained verbally

Test User	TP0 (Pilot Test)	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9
Alias	"Kim"	"Marshall"	"Teyana"	"Jermaine"	"Aubrey"	"Kendrick"	"Anthony"	"Arlissa"	"Gary"	"Robyn"
Date of Test	09/01/19	11/01/19	11/01/19	11/01/19	11/01/19	12/01/19	12/01/19	13.01.2019	13.01.2019	14.01.2019
Time of Test	05:45:00 PM	11:15:00 AM	01:50:00 PM	03:00:00 PM	05:20:00 PM	08:00:00 PM	08:45:00 PM	10:00:00 AM	11:00:00 AM	06:30:00 PM
Language of Test	EN	EN	EN	EN	EN	EN	EN	EN	EN	EN
General Information										
Sex	female	male	female	male	male	male	male	female	male	female
Age	23	23	22	19	22	22	22	20	22	20
Education	secondary school, studying interior design	secondary school, studying law	BA, International Management/ Economics	secondary school, studying CS	secondary school, studying law	secondary school, studying Software Engineering	secondary school, studying CS	secondary school, studying psychology	secondary school, studying law	secondary school, studying physiotherapy
Sight Impairment										
Sight Aid	glasses	glasses	glasses	-	-	-	-	glasses	glasses	glasses, contact lenses
Color Blindness	no	no	no	no	no	no	no	no	no	no
Personal Computer Use										
OS	Apple Macintosh	Apple Macintosh	Microsoft Windows	Microsoft Windows	Apple Macintosh	Microsoft Windows, Unix	Microsoft Windows	Apple Macintosh	Microsoft Windows	Microsoft Windows
Years of Experience	10	20	12	12	10	10	16	9	15	13
Hours per Week	5	50	35	55	25	30	60	14	10	20
Tablet Use										
Type of Tablet	-	-	iOS	iOS	-	-	iOS	-	-	iOS
Years of Experience	-	-	8	3	-	-	5	-	-	4
Hours per Week	-	-	20	3	-	-	1	-	-	2
Web Use										
Hours per Week	12	40	75	60	50	20	60	35	35	42
Most used Device	smart-phone	smart-phone	smart-phone	desktop PC	smart-phone	desktop PC	desktop PC	smart-phone	smart-phone	smart-phone
Browser	Safari	Chrome	Firefox	Chrome	Chrome Firefox	Firefox	Chrome	Safari	Chrome	Firefox
Domain-Specific Questions										
File-Managers used	Mac Finder	Windows Explorer, Mac Finder	Windows Explorer	Windows Explorer	Mac Finder	Windows Explorer	Windows Explorer	Mac Finder	Windows Explorer	Windows Explorer
Heard of Tree of Life	no	no	no	yes	no	yes	no	no	no	no
Browsed Product hierarchies	Amazon, Ebay, Clothing sites	Amazon, Ebay, Geizhals	Amazon	Amazon, Ebay, Willhaben	Amazon, Ebay, Willhaben, Shpock, Kleiderkreisel	Amazon, Ebay	Amazon, Ebay	Amazon	Amazon, Ebay, Geizhals	Amazon
Heard or seen Hyperbolic Tree Browser	no	no	no	no	no	no	no	no	no	no
Previous Usability Tests										
As Test Person	yes	yes	yes	yes	no	no	yes	no	no	no
In Test Team	no	no	no	no	no	yes	yes	no	no	-
Type of Test	Thinking Aloud	Formal Experiment	Formal Experiment	Thinking Aloud	-	Thinking Aloud	Thinking Aloud	-	-	-

Table 4.1: Overview of the 10 test users.

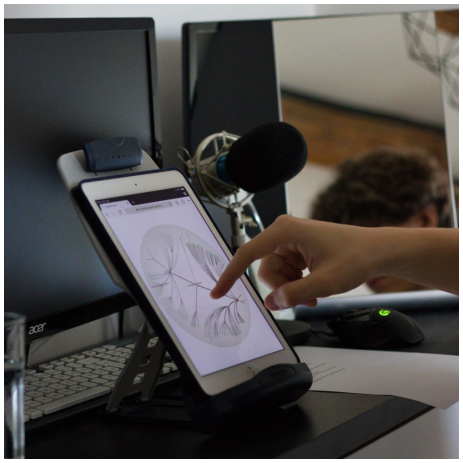


Figure 4.2: The tablet test setup [Photo taken by the author of this thesis, Christopher Oser].

Consent Form

Thank you for participating in our study. Please be aware that audio and video recordings will be made of your session, to allow others who are not present to observe your session and benefit from your feedback.

Please read the statement below and sign where indicated. Thank you.

I understand that audio and video recordings will be made of my session. I grant permission for these recordings to be used for teaching and research purposes.

Test User

Place: _____

Date: _____

Name: _____

Date of Birth: _____

Signature: _____

[Copyright © 2018 by Keith Andrews. This work is placed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence.]

Figure 4.3: The consent form presented to and signed by users [The template was provided by Keith Andrews].

to each test user and any questions were answered, but the technique itself was not practised. The test users were simply asked to follow the tasks and solve them one by one, whilst expressing their thoughts verbally in English.

4.6 Test Tasks

Each test was done in the same manner regardless of test user or device. After first contact the test users were given a consent form as well as a background information questionnaire that was to be filled out. The consent form can be seen in Figure 4.3. Once this had been done, the thinking aloud test was explained verbally to each subject in the same manner. The test users were asked to simply follow the tasks and solve them one by one all whilst expressing their thoughts verbally in English. The test users were able to read

Task	Description
1	Please navigate to the root of the Bear (Ursidae) subtree.
2	Find the node corresponding to the Brown Bear and read the first paragraph of the corresponding Wikipedia entry.
3	Would you say there are more species (sub-nodes) of Bear (Ursidae) or more of Dog (Canidae)?
4	Which species is more closely related to Whales (Baleen) according to this visualisation: Alpacas (Vicugna pacos) or Elefants (Elephantidae)?
5	Find the two black Howler Monkey nodes within the Primate tree.

Table 4.2: Tasks for the Tree of Life dataset.

Task	Description
1	Please navigate to and open up the Food branch of the Products tree.
2	Within this Food branch, please navigate to and open up the Donuts node.
3	Would you say there are more items in the Bakery branch (where you found Donuts) or in the Meat branch?
4	Please zoom back out to the root node. Which of these two items is within the Electronics branch: Antennas or Webcams?
5	Within the entire Products tree, please try to find the "Basketballs" node.

Table 4.3: Tasks for the Product Tree dataset.

the tasks right before starting the test and check them with the facilitator in order to prevent vocabulary misunderstandings. The tasks for each test can be found in Tables 4.2 and 4.3 respectively.

The presentation order of device and dataset was varied according to a pre-defined scheme. In this study, the pilot test with TP0 showed that the Tree of Life dataset was very hard to use on a tablet, so an executive decision was made to only test the Product Tree on the tablet for the remaining test users. The final presentation order for each test user is shown in Table 4.4.

After filling out the forms and being briefed about the procedure, the test user was guided to the desk where the device was set up and the tasks were laid out in front of them. They then proceeded to complete task after task until all of them were accomplished. During the tests, the facilitator stood behind the user taking notes as well as helping out in certain situations in order for the test to proceed smoothly.

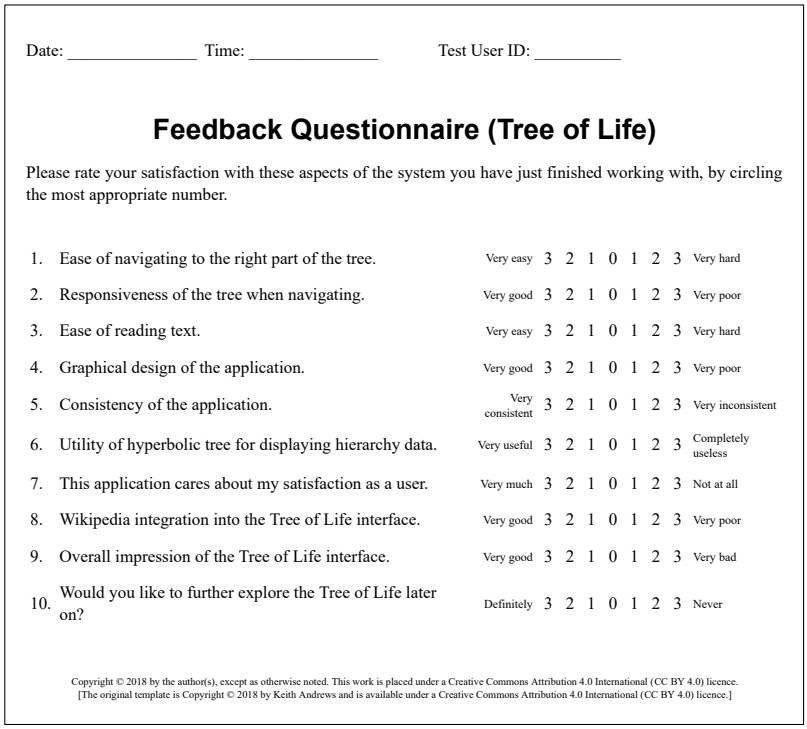
4.7 Interview

After completing the tasks, each user was asked questions like "How was it?" in order to capture their immediate reflections. The exact questions can be viewed below.

- Opening Question
  - "How was it?"
- Standard Questions
  - "Did anything strike you as particularly good?"
  - "Did anything strike you as particularly bad?"
  - "Do you have any other remarks, good or bad that come to mind?"
  - "Now after filling out the questionnaire did anything else come to mind you would want to say about the application?"

	First Session	Second Session
TP0	Tablet, Tree of Life	PC, Product Tree
TP1	Tablet, Product Tree	PC, Tree of Life
TP2	Tablet, Product Tree	PC, Tree of Life
TP3	Tablet, Product Tree	PC, Tree of Life
TP4	Tablet, Product Tree	PC, Tree of Life
TP5	PC, Tree of Life	Tablet, Product Tree
TP6	PC, Tree of Life	Tablet, Product Tree
TP7	PC, Tree of Life	Tablet, Product Tree
TP8	PC, Tree of Life	Tablet, Product Tree
TP9	PC, Tree of Life	Tablet, Product Tree

Table 4.4: The presentation order of devices and datasets.



Date: _____ Time: _____ Test User ID: _____

Feedback Questionnaire (Tree of Life)

Please rate your satisfaction with these aspects of the system you have just finished working with, by circling the most appropriate number.

1.	Ease of navigating to the right part of the tree.	Very easy	3	2	1	0	1	2	3 Very hard
2.	Responsiveness of the tree when navigating.	Very good	3	2	1	0	1	2	3 Very poor
3.	Ease of reading text.	Very easy	3	2	1	0	1	2	3 Very hard
4.	Graphical design of the application.	Very good	3	2	1	0	1	2	3 Very poor
5.	Consistency of the application.	Very consistent	3	2	1	0	1	2	3 Very inconsistent
6.	Utility of hyperbolic tree for displaying hierarchy data.	Very useful	3	2	1	0	1	2	3 Completely useless
7.	This application cares about my satisfaction as a user.	Very much	3	2	1	0	1	2	3 Not at all
8.	Wikipedia integration into the Tree of Life interface.	Very good	3	2	1	0	1	2	3 Very poor
9.	Overall impression of the Tree of Life interface.	Very good	3	2	1	0	1	2	3 Very bad
10.	Would you like to further explore the Tree of Life later on?	Definitely	3	2	1	0	1	2	3 Never

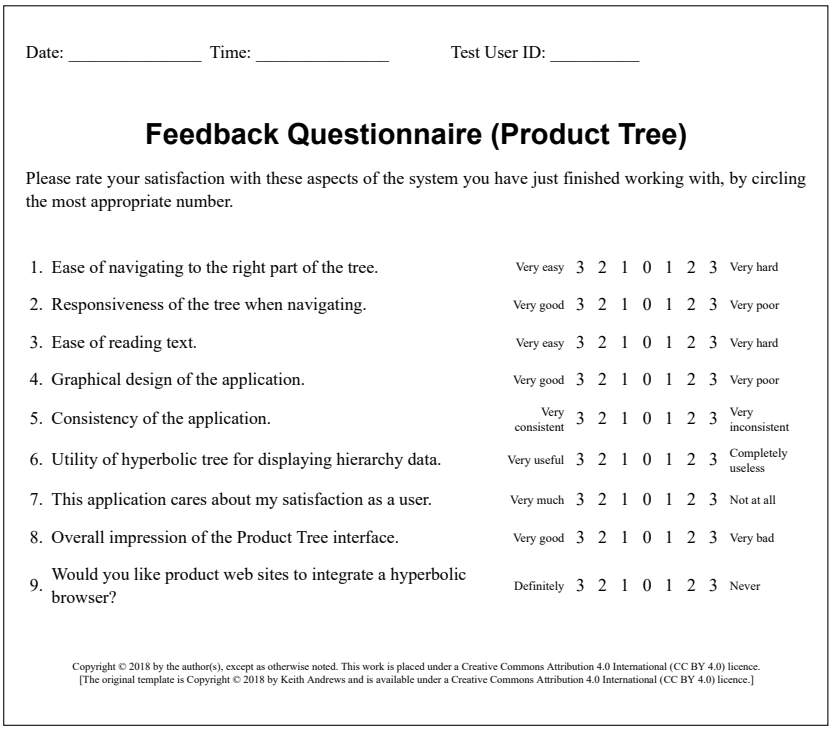
Figure 4.4: The feedback questionnaire presented to users after the Tree of Life dataset [The feedback questionnaire template was provided by Keith Andrews and adapted by Christopher Oser].

4.8 Feedback Questionnaire

After the interview, the test user was asked to complete the feedback questionnaire relevant to the dataset tested. There are two different feedback questionnaires used in the tests, one for each dataset. They can be found in Figures 4.4 and 4.5 respectively. If the user had no further remarks this concluded the test session.

4.9 Checklist

The whole procedure was done in the same manner for all tests and users. To prevent human error a checklist was used in order to keep the same procedure for all 20 test sessions. The checklist can be seen in Figure 4.6.



Date: _____ Time: _____ Test User ID: _____

Feedback Questionnaire (Product Tree)

Please rate your satisfaction with these aspects of the system you have just finished working with, by circling the most appropriate number.

1.	Ease of navigating to the right part of the tree.	Very easy	3	2	1	0	1	2	3 Very hard
2.	Responsiveness of the tree when navigating.	Very good	3	2	1	0	1	2	3 Very poor
3.	Ease of reading text.	Very easy	3	2	1	0	1	2	3 Very hard
4.	Graphical design of the application.	Very good	3	2	1	0	1	2	3 Very poor
5.	Consistency of the application.	Very consistent	3	2	1	0	1	2	3 Very inconsistent
6.	Utility of hyperbolic tree for displaying hierarchy data.	Very useful	3	2	1	0	1	2	3 Completely useless
7.	This application cares about my satisfaction as a user.	Very much	3	2	1	0	1	2	3 Not at all
8.	Overall impression of the Product Tree interface.	Very good	3	2	1	0	1	2	3 Very bad
9.	Would you like product web sites to integrate a hyperbolic browser?	Definitely	3	2	1	0	1	2	3 Never

Figure 4.5: The feedback questionnaire presented to users after the Product Tree dataset [The feedback questionnaire template was provided by Keith Andrews and adapted by Christopher Oser].

Checklist for Thinking Aloud Test

1. Preparation:

   ☐ Reset interface for new user.

   ☐ Check that everything is ready in test room.

2. Opening:

   ☐ Greet the participant.

   ☐ Ask user to read and sign consent and non-disclosure forms.

   ☐ Ask user to fill out Background Questionnaire and offer help.

   ☐ Explain the testing procedure.

3. Test Session:

   ☐ Move over to testing area.

   ☐ Double check with the user whether they understood the wording of the tasks.

   ☐ Start session recording.

   ☐ Start camera recording.

   ☐ User begins with tasks.

   ☐ User finishes last task.

4. Closing:

   ☐ Interview: how was it?

   ☐ Further interview questions.

   ☐ Individual interview questions arising from test.

   ☐ Feedback questionnaire. User fills out form.

   ☐ Thank participant, provide any remuneration, show participant out.

5. Wrap-Up:

   ☐ Summarise thoughts about this test.

   ☐ Organise data sheets and notes.

   ☐ Check and archive session recordings.

Figure 4.6: The checklist used to keep the same procedure during all 20 test sessions (two per test user). [The checklist template was provided by Keith Andrews and was adapted by Christopher Oser].

Chapter 5

Test Results

This chapter documents the results of the thinking aloud test of the d3-hypertree hierarchy browser. With each test, it became more clear what the main negative points to average users are. Almost every test user had the same issues, stemming from a mixture of performance problems and design decisions or missing features. In terms of performance, one of the main issues is the performance on a tablet, which was frequently laggy and would sometimes freeze. Another major problem were crashes of the tree when performing rash movements or zooming.

Design-wise, a pressing problem is the implementation of Wikipedia synchronized to the Tree of Life on tablets, which led to the cancellation of testing of the Tree of Life on tablets altogether. The Wikipedia page does not show up at all on tablets. It needs to be drawn out from the side of the screen, which in consequence makes the rest of the tree too small to be used. The main design problems were overview and search capabilities. All users agreed that the overview provided by the d3-hypertree was mediocre at best. It is very easy to become lost and to not arrive at the desired location even if it has been explored before. Some users proposed a map or working with colors to make it easier to distinguish between locations.

Searching for a specific tree or node without knowing its location in advance is very slow compared to text search. This is why most users proposed a search bar that would be linked to the nodes and would enable users to find nodes by entering a search term. Once a certain node is found, the d3-hypertree is perfect for browsing the similar categories and much more intuitive than other visualizations for similar datasets.

5.1 Task Completion

No test user failed to complete a task. Nonetheless, every task had been previously assigned a time limit, which indicated whether the test user was able to complete the task in a reasonable time frame. If the time limit was exceeded, the attempt counted as failed. A summary of task completion for each test and test user can be viewed in Tables 5.1 and 5.2. If the user was truly stuck, the facilitator provided some assistance or hints, so the user could proceed.

5.2 Positive Findings

While the main focus is to find negative issues with the application during testing in order to fix them, the positives are also noted to give the developers some positive feedback too. The three most positive findings according to their positivity ratings are described below. All three are illustrated in Figure 5.1. The positivity rating scheme used to rank positive findings is shown in Table 5.3. All further positive findings, and their ratings are listed in Table 5.4.

	Task 1	Task 2	Task 3	Task 4	Task 5
TP0	1	0	1	1	0*
TP1	1	0	1	1	0*
TP2	0*	1	1	1	1
TP3	0	1	1	1	1
TP4	1	1	1	1	1
TP5	1	1	1*	1	1
TP6	0	1	1	1	1
TP7	1	1	1	1	1*
TP8	1	1	1	0	1
TP9	1	1	1*	1	0
Total	7	8	10	9	7
%	70	80	100	90	70

Table 5.1: Task completion rates for the Tree of Life. An asterisk (*) indicates that assistance was given.

	Task 1	Task 2	Task 3	Task 4	Task 5
TP0	1	1	1	1	1
TP1	1	1	1	1	1
TP2	1	1	1	1	1
TP3	1	1	1	1	1
TP4	1	1	1	1	1
TP5	1	1	1*	1	1
TP6	1	1	1	0	1
TP7	1	1	1	1	1
TP8	1	1	1	1	1
TP9	1	1	1	1	0
Total	10	10	10	9	9
%	100	100	100	90	90

Table 5.2: Task completion rates for the Product Tree. An asterisk (*) indicates that assistance was given.

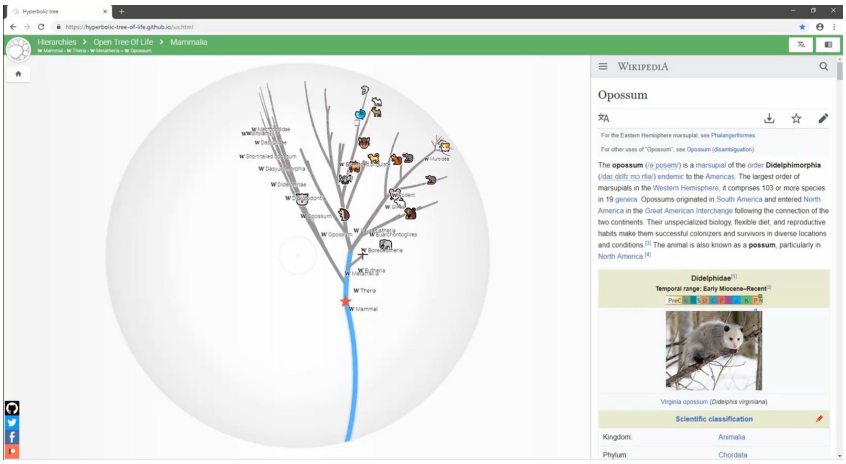


Figure 5.1: The screenshot shows all of the positive findings: the icons/emojis, the Wikipedia integration, and the home star icon. [The screenshot was taken by the author of this thesis, Christopher Oser]

Positivity	Meaning
4	Extremely Positive
3	Major Positive
2	Minor Positive
1	Cosmetic Positive
0	Not a Positive

Table 5.3: Positivity ratings used to rank the positive findings.

5.2.1 Positive 1: Icons/Emoji

Most test users very much appreciated the emojis and said they would be lost without them. Three test users specifically requested more icons and/or more use of colors or other functions in order to increase the distinguishability, since the tree can become rather overwhelming without certain landmarks or markers.

5.2.2 Positive 2: Wikipedia Integration

Most users were very impressed by the synchronization with Wikipedia in the Tree of Life dataset. Some used it in order to fulfill tasks although not requested to. The Wikipedia part was used far more than expected, as users are used to it and found all kinds of uses for it. Some requested that there would be a link from Wikipedia pages to the respective nodes. This would be one way to resolve the lack of a search function described in Section 5.3.3.

5.2.3 Positive 3: Home Star Icon

The home star icon provided by d3-hypertree was a very useful visual landmark to test users. It seems users are thankful for any type of distinguishability or markers that guide them through these large tree datasets.

5.2.4 All Positive Findings

Table 5.4 lists all of the positive findings which emerged from the test, sorted in decreasing order of positivity rating, so the most positive are at the top of the table. Positives 0 and 6 were found during the post-test interviews.

5.3 Negative Findings

The main reason for testing an application is usually to find problems which have yet to be considered during development. When testing d3-hypertree the main problems were clear early on since most test users ran into similar issues. The top five problems according to their severity ratings are described below. The severity rating scheme used to rank the problems is shown in Table 5.5. All problems are listed in Table 5.6.

5.3.1 Problem 1: Lack of Overview

Very often, users were overwhelmed by the situation and were lacking an overview over the entire tree. Things like a "mini-map" or some other way to create more overview would be helpful. This problem can be seen in Figure 5.2.

No.	Title	Description	Timestamps	Location (how reproducible?)	Positivity
0	Concept	9 out of 10 test users liked the concept and could see it being used in major applications.	Post-test interview	d3-hypertree.	4
1	Icons/Emoji	Most users very much appreciated the emojis.	tp1 ToL 00:38, tp2 ToL 02:30	Tree of Life d3-hypertree.	4
2	Wikipedia Integration	Users very impressed by the integration of Wikipedia.	tp1 ToL 14:10, tp2 ToL 03:30, tp6 ToL 03:30	Tree of Life d3-hypertree.	4
3	Home Star Icon	The root node marker represented by a star was very helpful to test users.	tp0 Prod 03:20, tp2 Prod 02:22, tp2 Prod 03:35	From any point within the tested datasets, a star iocn can be seen which takes you back to the root node.	3
4	Consistency	Users said the second test is much easier. After a short steep learning curve, the application seems to become easier to use.	tp0 ToL 00:28, tp1 Prod 00:26, tp6 Prod 05:15	Use the d3-hypertree for a while.	2
5	Alphabetical Ordering	Some users were very happy with the alphabetically ordered data in the product tree.	tp2 Prod 03:50	The labels of the Product Tree.	1
6	360° Layout	2 users greatly preferred the 360° layout of the Product Tree over the more "tree-like" view in the Tree of Life.	Post-test interview	Product Tree d3-hypertree.	1

Table 5.4: List of all positive findings, in descending order of positivity.

Severity	Meaning
4	Catastrophic problem
3	Serious problem
2	Minor problem
1	Cosmetic problem
0	Not a problem

Table 5.5: Severity ratings used to rank the problems found.

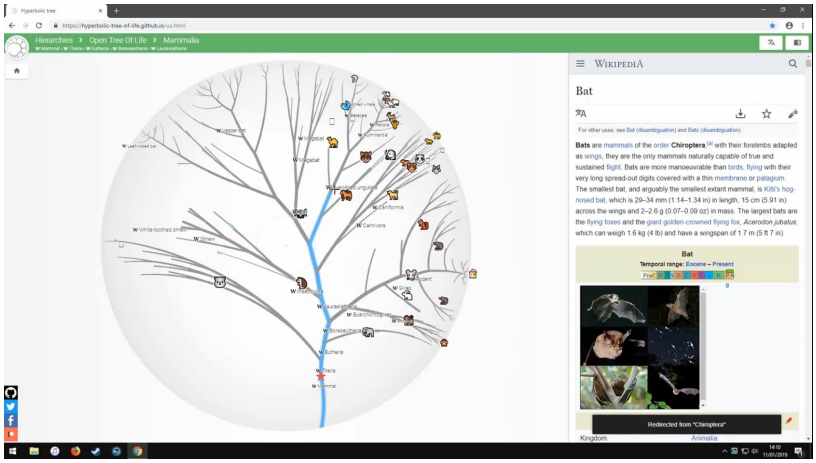


Figure 5.2: The lack of overview described by Problem 1. [Screenshot taken by the author of this thesis, Christopher Oser]

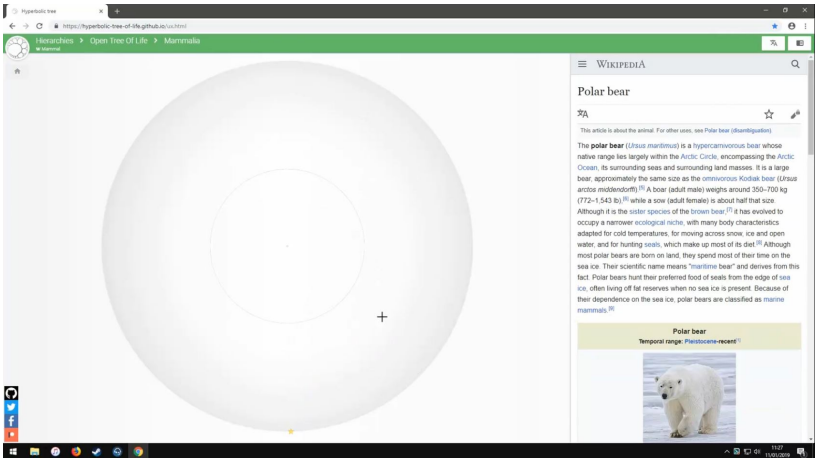


Figure 5.3: One of the many crashes experoenced by test users. [Screenshot taken by the author of this thesis, Christopher Oser]

5.3.2 Problem 2: Random Crashes

Sudden crashes are very annoying, especially if they happen as often as during the test sessions. Some users managed to crash the tree three or four times within one session, which would make regular users stop using it. This problem can be seen in Figure 5.3.

5.3.3 Problem 3: Lack of Search Function

Most tasks took a long time, due to the lack of knowledge about the dataset as well as the overwhelming nature of the tree. A simple text search function would make searching for a specific item much quicker and more intuitive for users, who can then proceed to explore further from the node they searched for.

5.3.4 Problem 4: Tablet Performance

Although tasks could still be accomplished the performance on a tablet was very slow and annoying to users. It sometimes takes more than a second for a user action to be executed on the tree.

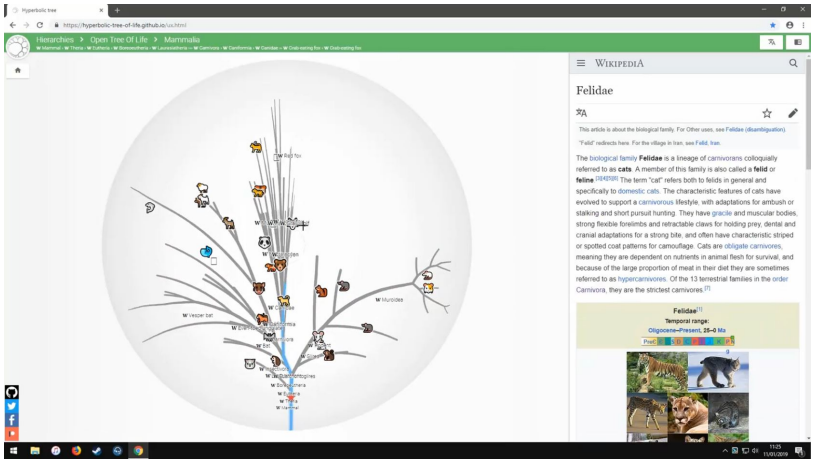


Figure 5.4: The interface lacks a tutorial or introduction. [Screenshot taken by the author of this thesis, Christopher Oser]

5.3.5 Problem 5: Lack of Tutorial/Introduction

Most users were overwhelmed for a while when first using the d3-hypertree. A built-in introduction to the entire concept as well as little things like a zoom indicator or similar helpers would ameliorate this. This problem can be seen in Figure 5.4.

5.3.6 All Problems Found

Table 5.6 lists all the problems observed in the test, sorted in decreasing order of severity, meaning the most severe are at the top of the table. The rating scale used for these findings can be found in Table 5.5. Problem No.9 was found in post-test interviews.

5.4 Interviews

Nine out of ten test users liked the concept of hyperbolic browsing and the d3-hypertree and could imagine using this somewhere else more regularly. Only one user was strongly opposed to the idea of using this in everyday situations and thought it was impracticable. Seven users requested more distinguishability and every single one said more overview is needed. Since there is so much data on the screen, it can become overwhelming to users very quickly.

Four users did say it would need some more work in order to be implemented into existing platforms, as it is not that intuitive to an average user. The utility of the Tree of Life was confirmed by all of the test users. Four test users said the Product Tree needed some adaptations in order to be used in existing product taxonomies. Although three users also mentioned it being more neatly arranged, due to its alphabetical ordering and the 360° spread of the tree.

5.5 Feedback Questionnaire

Table 5.7 shows the ratings given by users in the feedback questionnaire at the end of the Tree of Life test. Table 5.8 shows the ratings given by users in the feedback questionnaire at the end of the Product Tree test. The neutral scale in the original feedback questionnaires has been mapped to a point scale between 6 (best) and 0 (worst). The numbers in bold indicate the (rounded) mean rating. The original blank questionnaires can be seen in Figures 4.4 and 4.5.

No.	Title	Description	Timestamps	Location (how reproducible?)	Severity
1	Lack of Overview	The d3-hypertree is very overwhelming for new users and lacks any form of overview.	tp0 ToL 05:30, tp1 ToL 05:40, tp2 ToL 09:08, . . .	Using d3-hypertree.	4
2	Random Crashes	The d3-hypertree tends to crash on both PC and tablet (although more on PC) when zooming too aggressively.	tp0 ToL 06:35, tp1 ToL 04:00, tp2 ToL 05:30, . . .	Zooming in or out or moving around very fast.	4
3	Lack of Search Function	It is very hard to find things within the tree without any prior knowledge.	tp0 Prod 04:00, tp0 ToL 07:25, tp1 ToL 07:35, . . .	Finding something in d3-hypertree.	4
4	Tablet Performance	The performance of the application is very slow on tablet.	tp0 ToL 04:50, tp1 Prod 01:00, tp2 Prod 00:15, . . .	Using d3-hypertree on tablet.	3
5	Lack of Tutorial/Introduction	Users ran into problems, due to not knowing how to interact with the tree or how it is set up.	tp0 ToL 09:30, tp1 ToL 02:10, tp2 ToL 02:33, . . .	Initial use of d3-hypertree.	3
6	Tablet Wikipedia	Wikipedia integration is very unusable on tablet. It does not show at all at first. If it shows, it messes with the rest of the page.	tp0 ToL 02:00	Using Tree of Life on tablet and trying to use Wikipedia integration	3
7	Disappearing Labels	Labels tend to disappear randomly or not show at all unless explicitly zoomed into.	tp0 Prod 05:00, tp2 Prod 01:35, tp6 Prod 00:30, . . .	Zoom around the d3-hypertree.	2
8	Data Inconsistency	Data inconsistencies (2 Meat branches in Product Tree) are very unsettling to users.	tp1 Prod 02:20, tp2 Prod 03:10	Path to meat branches in Product Tree	2
9	Monkey Emojis Misleading	Within the Primate tree, 3 different monkey emojis (closed mouth, closed eyes, closed ears) were used, which rather confused users.	Post-test interview	Monkey node in the Tree of Life.	1
10	Infinite Zoom	It can happen rather easily on tablet that the user zooms to the far end of the globe, which makes it seem like the application crashed.	tp1 Prod 04:40	Quick gestures on tablet.	1

Table 5.6: List of all problems found, in descending order of severity.

					TP0	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	Mean	Std Dev
1	Ease of navigating to the right part of the tree.	Very easy	6 5 **4** 3 2 1 0	Very hard	1	1	4	6	5	6	5	4	6	5	4.3	1.89
2	Responsiveness of the tree when navigating.	Very good	6 **5** 4 3 2 1 0	Very poor	1	6	5	5	5	6	6	5	4	6	4.9	1.52
3	Ease of reading text.	Very easy	6 **5** 4 3 2 1 0	Very hard	5	5	5	5	3	6	4	6	6	6	5.1	0.99
4	Graphical design of the application.	Very satisfied	6 **5** 4 3 2 1 0	Very unsatisfied	5	5	5	6	2	5	4	5	5	6	4.8	1.14
5	Consistency of the application.	Very satisfied	**6** 5 4 3 2 1 0	Very unsatisfied	6	4	6	6	5	6	5	6	5	6	5.5	0.71
6	Utility of the hyperbolic tree for displaying hierarchy data.	Very satisfied	6 **5** 4 3 2 1 0	Very unsatisfied	5	4	5	5	2	6	3	6	6	5	4.7	1.34
7	This application cares about my satisfaction as a user.	Very satisfied	6 5 4 **3** 2 1 0	Very unsatisfied	1	1	1	5	1	5	4	5	5	5	3.3	2.00
8	Wikipedia integration into the Tree of Life interface.	Very satisfied	6 **5** 4 3 2 1 0	Very unsatisfied	0	5	6	6	6	5	6	6	6	6	5.2	1.87
9	Overall impression of the Tree of Life interface.	Definitely	6 5 **4** 3 2 1 0	Never	3	4	4	6	4	4	3	5	6	5	4.4	1.07
10	Would you like to further explore the Tree of Life later on?	Very satisfied	6 **5** 4 3 2 1 0	Very unsatisfied	4	5	3	6	1	5	4	6	6	5	4.5	1.58

Table 5.7: User ratings from the Tree of Life feedback questionnaire.

					TP0	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	Mean	Std Dev
1	Ease of navigating to the right part of the tree.	Very easy	6 5 **4** 3 2 1 0	Very hard	2	5	6	6	2	6	1	6	5	5	4.4	1.96
2	Responsiveness of the tree when navigating.	Very good	6 5 **4** 3 2 1 0	Very poor	6	4	4	4	4	4	1	6	2	1	3.6	1.78
3	Ease of reading text.	Very easy	6 **5** 4 3 2 1 0	Very hard	4	5	5	6	1	2	5	6	6	6	4.6	1.78
4	Graphical design of the application.	Very satisfied	6 **5** 4 3 2 1 0	Very unsatisfied	4	5	5	6	2	5	5	5	5	5	4.7	1.06
5	Consistency of the application.	Very satisfied	**6** 5 4 3 2 1 0	Very unsatisfied	6	5	6	6	5	6	6	6	5	5	5.6	0.52
6	Utility of the hyperbolic tree for displaying hierarchy data.	Very satisfied	6 **5** 4 3 2 1 0	Very unsatisfied	5	5	6	6	1	6	5	4	4	5	4.7	1.49
7	This application cares about my satisfaction as a user.	Very satisfied	6 5 **4** 3 2 1 0	Very unsatisfied	4	3	5	5	3	5	3	5	3	6	4.2	1.14
8	Overall impression of the Product Tree interface.	Very satisfied	6 **5** 4 3 2 1 0	Very unsatisfied	5	5	5	6	2	5	4	6	5	5	4.8	1.14
9	Would you like product web sites to integrate a hyperbolic browser?	Definitely	6 5 **4** 3 2 1 0	Never	5	4	6	5	0	5	1	5	5	5	4.1	1.97

Table 5.8: User ratings from the Product Tree feedback questionnaire.

Chapter 6

Concluding Remarks

This thesis presented a thinking aloud usability test of the d3-hypertree web application for visualizing large hierarchies. Chapter 2 describes the thinking aloud test methodology and history. The ideas behind hyperbolic browsing for large hierarchies are presented in Chapter 3. Chapter 4 describes the test procedure and Chapter 5 presents the results of the test.

The results of the test highlighted a number of issues. Most test users requested more overview and distinguishability in the form of an overview map, color-coding, or landmark icons. The application also crashed quite frequently during the tests, which caused some consternation among the test users. To help locate specific items in the tree, many users requested a text search function. Nonetheless, the majority of test users could see the utility of the application in everyday life situations and would not be surprised to see it be used commercially in the future.

The online version of the test report [Oser 2019] includes video clips for each finding and transcripts of each test session.

Bibliography

Andrews, Keith [2018]. *Human-Computer Interaction: Course Notes*. 2018. https://courses.isds.tugraz.at/hci/hci.pdf (cited on page 3).

Andrews, Keith [2019]. *Writing a Thesis: Guidelines for Writing a Master's Thesis in Computer Science*. Graz University of Technology, Austria. 24 Jan 2019. http://ftp.isds.tugraz.at/pub/keith/thesis/ (cited on page ix).

Barnum, Carol M. [2010]. *Usability Testing Essentials*. Elsevier, 2010. ISBN 012375092X (cited on page 3).

Borsci, Simone and Stefano Federici [2009]. *The Partial Concurrent Thinking Aloud: A New Usability Evaluation Technique for Blind Users*. Assistive Technology from Adapted Equipment to Inclusive Environments—AAATE 25 (2009), pages 421–425. doi:10.3233/978-1-60750-042-1-421 (cited on page 3).

Cooke, Lynne [2010]. *Assessing Concurrent Think-Aloud Protocol as a Usability Test Method: A Technical Communication Approach*. IEEE Transactions on Professional Communication 53.3 (2010), pages 202–215. doi:10.1109/TPC.2010.2064812 (cited on page 3).

Ericsson, K. Anders and Herbert A. Simon [1980]. *Verbal Reports as Data*. Psychological Review 87.3 (1980), pages 215–251. doi:10.1037/0033-295X.87.3.215 (cited on page 3).

Ericsson, K. Anders and Herbert A. Simon [1984]. *Protocol Analysis: Verbal Reports as Data*. MIT Press, 1984. ISBN 0262050471 (cited on page 3).

Ericsson, K. Anders and Herbert A. Simon [1987]. *Verbal Reports on Thinking* (1987) (cited on page 3).

Federici, Stefano, Simone Borsci, and Gianluca Stamerra [2010]. *Web Usability Evaluation with Screen Reader Users: Implementation of the Partial Concurrent Thinking Aloud Technique*. Cognitive processing 11.3 (01 Aug 2010), pages 263–272. ISSN 1612-4782. doi:10.1007/s10339-009-0347-y (cited on page 3).

Forder, Henry George [1958]. *The Foundations of Euclidean Geometry*. Dover Publications, 1958 (cited on page 5).

Glatzhofer, Michael [2018a]. *d3-hypertree Mammal-Tree*. 2018. https://hyperbolic-tree-of-life.github.io/ux.html (cited on page 7).

Glatzhofer, Michael [2018b]. *Hyperbolic Tree of Life*. 2018. https://hyperbolic-tree-of-life.github.io/ (cited on page 9).

Glatzhofer, Michael [2019a]. *d3-hypertree Source Code*. 2019. https://github.com/glouwa/d3-hypertree (cited on page 5).

Glatzhofer, Michael [2019b]. *Hyperbolic Browsing: Scalable Hierarchy Browsing in Hyperbolic Space*. Draft Seminar Paper, Graz University of Technology. 2019 (cited on pages IX, XI, 5).

Google [2015]. *Google Product Taxonomy*. 19 Feb 2015. https://google.com/basepages/producttype/taxonomy.en-US.txt (cited on pages 7–9).

Hinchliff, Cody E., Stephen A. Smith, James F. Allman, J. Gordon Burleigh, Ruchi Chaudhary, Lyndon M. Coghill, Keith A. Crandall, Jiabin Deng, Bryan T. Drew, Romina Gazis, Karl Gude, David S. Hibbett, Laura A. Katz, H. Dail Laughinghouse, Emily Jane McTavish, Peter E. Midford, Christopher L. Owen, Richard H. Ree, Jonathan A. Rees, Douglas E. Soltis, Tiffani Williams, and Karen A. Cranston [2015]. *Synthesis of Phylogeny and Taxonomy into a Comprehensive Tree of Life*. Proceedings of the National Academy of Sciences 112.41 (2015), pages 12764–12769. ISSN 0027-8424. doi:10.1073/pnas.1423041112. https://www.pnas.org/content/112/41/12764.full.pdf (cited on pages 5, 9).

Krahmer, Emiel and Nicole Ummelen [2004]. *Thinking About Thinking Aloud: A Comparison of Two Verbal Protocols for Usability Testing*. IEEE Transactions on Professional Communication 47.2 (2004), pages 105–117. doi:10.1109/TPC.2004.831549 (cited on page 3).

Lamping, John and Ramana Rao [1994]. *Laying Out and Visualizing Large Trees Using a Hyperbolic Space*. Proc. 7th Annual ACM Symposium on User Interface Software and Technology (UIST '94) (Marina del Rey, United States). ACM. 1994, pages 13–14. doi:10.1145/192426.192430 (cited on page 5).

Lamping, John and Ramana Rao [1996]. *Visualizing Large Trees Using the Hyperbolic Browser*. Proc. Conference on Human Factors in Computing Systems (CHI '96) (Vancouver, Canada). ACM. 1996, pages 388–389. doi:10.1145/257089.257389 (cited on page 5).

Lamping, John, Ramana Rao, and Peter Pirolli [1995]. *A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies*. Proc. Conference on Human Factors in Computing Systems (CHI '95) (Denver, Colorado). ACM. 1995, pages 401–408. doi:10.1145/223904.223956 (cited on page 5).

Lewis, Clayton H. [1982]. *Using the "Thinking-Aloud" Method in Cognitive Interface Design*. Research Report RC 9265. IBM TJ Watson Research Center, 17 Feb 1982. https://domino.research.ibm.com/library/cyberdig.nsf/3addb4b88e7a231f85256b3600727773/2513e349e05372cc852574ec0051eea4 (cited on page 3).

Lewis, Clayton and John Rieman [1993]. *Task-Centered User Interface Design*. A practical introduction (1993) (cited on page 3).

Munzner, Tamara [1998a]. *Drawing Large Graphs With H3viewer and Site Manager*. Proc. International Symposium on Graph Drawing (GD 1998) (Montréal, Canada). Springer, 13 Aug 1998, pages 384–393. doi:10.1007/3-540-37623-2_30 (cited on page 5).

Munzner, Tamara [1998b]. *Exploring Large Graphs in 3D Hyperbolic Space*. IEEE Computer Graphics and Applications 18.4 (1998), pages 18–23. ISSN 0272-1716. doi:10.1109/38.689657 (cited on page 5).

Nielsen, Jakob [1994a]. *Estimating the Number of Subjects Needed for a Thinking Aloud Test*. International Journal of Human-Computer Studies 41.3 (Sep 1994), pages 385–397. doi:10.1006/ijhc.1994.1065 (cited on page 3).

Nielsen, Jakob [1994b]. *Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier*. In: *Cost-Justifying Usability*. Edited by Randolph G. Bias and Deborah J. Mayhew. 1st Edition. Academic Press, 1994. Chapter 11, pages 245–272. ISBN 0120958104 (cited on page 3).

Nielsen, Janni, Torkil Clemmensen, and Carsten Yssing [2002]. *Getting Access to What Goes On in People's Heads?: Reflections on the Think-Aloud Technique*. Proc. Nordic Conference on Human-Computer Interaction (NordiCHI 2002) (Aarhus, Denmark). ACM. 19 Oct 2002, pages 101–110. doi:10.1145/572020.572033 (cited on page 3).

Oser, Christopher [2019]. *Thinking Aloud Test Report - Usability Testing d3-hypertree Hierarchy Browser*. Feb 2019. https://projects.isds.tugraz.at/hypertree/ta-2019-01/ta.html (cited on page 25).

Oser, Christopher and Michael Glatzhofer [2018]. *d3-hypertree Product Tree*. 2018. https://b-lovedconqueror.github.io/ (cited on pages 8–9).

Rees, Jonathan A. and Karen Cranston [2017]. *Automated Assembly of a Reference Taxonomy for Phylogenetic Data Synthesis*. Biodiversity Data Journal 5 (2017), e12581. ISSN 1314-2836. doi:10.3897/BDJ.5.e12581 (cited on pages 5, 9).

Stillwell, John [1996]. *Sources of Hyperbolic Geometry*. 10. American Mathematical Society, 1996. ISBN 0821809229 (cited on page 5).

Van Den Haak, Maaike, Menno De Jong, and Peter Jan Schellens [2003]. *Retrospective vs. Concurrent Think-Aloud Protocols: Testing the Usability of an Online Library Catalogue*. Behaviour & Information Technology 22.5 (2003), pages 339–351. doi:10.1080/01449290310001632401 (cited on page 3).