

**Kandiatnr 305**

---

# **To Bento or Not to Bento?**

**A Usability Study of Prototypes of Oria and Bento Box**

**Master thesis 2020  
Master in Library and Information Science  
OsloMet – Oslo Metropolitan University,  
Department of Archivistcs, Library and Information Sciences**

## Abstract

### English

This study focuses on a usability test conducted on two prototypes of the websites — Norwegian Primo (Oria of the University of Oslo Library) and a Bento box (the Princeton University Library). I conducted a virtual face-to-face usability test using Zoom video meeting, with pre- and after- interviews to collect data. Eight undergraduate students participated one by one. The participants' interactions with the prototypes were observed via Zoom "shared screen".

The purpose of this study is to investigate how academic library users experienced using a Bento box interface and to identify ways to improve Oria. Results indicated that users had a strong preference for Bento box over Oria; participants were satisfied with Oria's user-friendly design and effectiveness. They liked a single search box and struggled with using faceted filters, while recognizing that faceted filters were more effective and discoverable. I also identified two major usability problems in terminology and facets menu design. The information collected about website navigation and searching practices provide insight into students' search habits using discovery tools like Primo and experience using Bento box. Lessons learned may assist other organizations in developing better searching tools.

### Norsk

I denne masteroppgaven sammenlignes prototyper av bruker-grensesnitt til to fagbiblioteks søketjenester (Oria med fasettert søk og Bento box). Åtte bachelor studenter deltok individuelt i bruker-test undersøkelse via video møte på Zoom. Observasjonen av deltakernes søkeprosess via Zoom "Del skjermen" funksjon. Data samles via før og etter intervjuer og observasjoner.

Målet er å vise hvordan brukere i fagbibliotek opplevde å bruke Bento box søketjeneste og identifisere muligheter for å forbedre Oria. Resultatene avdekket at brukerne fortrakk Bento box grensesnittet, samt at de anerkjente at fasettert søk var mer effektivt og nyttig. Studien identifiserte også to viktige problemer i terminologi og fasettmenydesign. Data samlet om navigasjons på nettstedet ga innsikt i studentenes informasjons-søkeatferd ved å bruke Primo og opplevelse av å bruke Bento box. Disse erfaringer kan hjelpe andre organisasjoner med å utvikle bedre søketjenester.

## **Preface**

Two-year master studies for me was an unexpected choice in my life. I had never thought I would choose any other subjects than literature and language. But after I finished my bachelor studies in library and information science, I found it triggered my interest in this subject. Two-year was short, but I have learned a lot. Library and information science shed a light on a new area in my life. Like the topic in this thesis, I spent over one year indulging myself into a new field – usability studies. I have read books and articles about usability studies about discovery tools and library websites, card sorting, paper prototyping and webpage design. I have learned how to make prototypes, use online video testing tools, and so on. For me, writing this thesis was an interesting process of studying.

I want to acknowledge the encouragement from both Oslo Metropolitan University and University of Oslo, where professors, librarians and classmates gave me both support and suggestions to conduct this test. I would show my appreciation to all of you, thank for sharing knowledge and friendships with me. I will especially thank my supervisor Gerd Berget for providing invaluable advice for this project: with refining the research question, from test design to thesis writing. Pia Borlund inspired me to start with usability testing when I followed her course. Dan Michael, O. Heggø and Line Nybakk Akerholt from University of Oslo Library gave me the idea to conduct a usability study about Oria and Bento box. Thank you to all the participants for your dedication and your patience during the test, especially in the middle of the corona pandemic. Without you, there would be no thesis.

Oslo, June 2020

# Table of Contents

<b>1. Introduction .....</b>	<b>3</b>
<b>1.1 Background .....</b>	<b>3</b>
1.1.1 Primo –One of the Most Popular Discovery Tools in Academic Libraries .....	4
1.1.2 Bento box – One for All? .....	5
<b>1.2 Motivation for This Study .....</b>	<b>7</b>
<b>1.3 Research Question and Chapter Overview .....</b>	<b>8</b>
<b>2. Theory and Literature Review .....</b>	<b>10</b>
<b>2.1 IIR and Information Search Behavior.....</b>	<b>10</b>
2.1.1 Three Information Retrieval Approaches.....	10
2.1.2 Information Searching Behavior and Information Seeking Behavior Models.....	11
<b>2.2 Usability Testing .....</b>	<b>13</b>
2.2.1 Factors to Measure Usability.....	13
2.2.2 Users and Goals.....	14
2.2.3 Profile of Target Audience .....	14
2.2.4 User Tasks Setting Scenarios .....	15
<b>2.3 Literature Review .....</b>	<b>15</b>
2.3.1 User Studies of Discovery Tools.....	15
2.3.2 Bento Box Search Interface Studies.....	17
<b>3. Methodology .....</b>	<b>20</b>
<b>3.1 Usability Study Methods Evaluation.....</b>	<b>20</b>
3.1.1 The Benefits and Limitations of the Methods I did not Choose.....	21
3.1.2 Card Sorting or Not? .....	23
3.1.3 Wireframes and Paper Prototype.....	25
3.1.4 Advantages and Limits of the Chosen Methods.....	25
3.1.5 Do- It- Yourself Testing .....	26
<b>3.2 Design Theory and Test Theory in Practice.....</b>	<b>27</b>
3.2.1 Simulated Work Task Situation and Tasks Setting .....	27
3.2.2 Preference VS. Performance .....	31
3.2.3 “Think-aloud” Current VS. “Think-aloud” Retrospective Methodology.....	32
3.2.4 Balsamiq Cloud Wireframes and Zoom.....	33
<b>3.3 Ethical Considerations .....</b>	<b>33</b>
<b>4. Test Process.....</b>	<b>35</b>
<b>4.1 Tasks Design, Pre- and After- Interviews.....</b>	<b>35</b>
<b>4.2 Making Wireframes, Video Observation and Recruiting Participants .....</b>	<b>38</b>
4.2.1 Making Wireframes.....	38
4.2.2 Video Remote Observation .....	46
4.2.3 Recruiting Participants .....	47
<b>4.3 Pilot Test .....</b>	<b>48</b>
<b>4.4 Observation and Interview via Video Online Test .....</b>	<b>49</b>
<b>5. Results and Analysis .....</b>	<b>51</b>
<b>5.1 Who are Test Participants? .....</b>	<b>51</b>

5.2 Information Search Behavior When Using Bento Box.....	53
5.3 Information Search Behavior When Using Oria .....	59
5.4 Performance of Searching and Tasks Understanding.....	64
5.5 Preference for Bento Box or Oria .....	67
5.6 Undergraduates’ Information Competency: Understanding Facets and Library Terminology .....	69
5.7 Findings: To Bento or Not to Bento? .....	70
5.7.1 Benefits and Disadvantages of “To Bento” .....	72
5.7.2 Benefits and Disadvantages of “Not to Bento” .....	73
5.7.3 Reality and Expectation.....	73
<b>6. Conclusions, Limitations and Further Study Suggestions .....</b>	<b>75</b>
6.1 Conclusions.....	75
6.2 Limitations.....	77
6.3 Contribution and Further Study Suggestions .....	78
6.3.1 Recommendations for Further Usability Test .....	78
6.3.2 Card Sorting.....	80
6.3.3 Students’ Information Competency and Library Information Literacy Lectures .....	81
<b>References: .....</b>	<b>82</b>
<b>Appendix .....</b>	<b>87</b>
<b>A1 Invitation letter .....</b>	<b>87</b>
<b>A2 Think Aloud Form.....</b>	<b>88</b>
<b>A3 Test Introduction .....</b>	<b>90</b>
<b>A4 Pre –interview Guide.....</b>	<b>91</b>
<b>A5 After-interview Guide .....</b>	<b>92</b>
<b>List of Figures</b>	
Figure 1 Three Approaches to Information Retrieval System Evaluation .....	11
Figure 3 Norman’s Gulfs Model .....	12
Figure 4 Single Search Box in Bento Box .....	39
Figure 5 Title Search in Bento Box.....	39
Figure 6 Author Search in Bento Box .....	39
Figure 7 Key Words Search in Bento Box .....	39
Figure 8 Oria Search box.....	39
Figure 9 Title Search in Oria .....	39
Figure 10 Title Search Results in Bento box.....	40
Figure 11 Author Search Results in Bento box .....	41
Figure 12 Key Words Search Results in Bento Box .....	42
Figure 13 Title Search Results in Oria .....	43
Figure 14 Title Search Results with Facet ‘E-book’ in Oria.....	44
Figure 15 Final Search Result in Oria .....	45
Figure 16 Aggregate Performance of Tasks in Bento Box .....	56

Figure 17 Content Type Facet Option Unobtrusive Button “Apply” in Bento box .....	57
Figure 18 Content Type Facet Option Button “Apply” in Bento box .....	58
Figure 19 Invisible Button on the Top of Facets Menu in Oria .....	61
Figure 20 Invisible Drop Down Menu on the Top of Facets Menu in Oria.....	61
Figure 21 Aggregate Performance of Tasks in Oria.....	63
Figure 22 Comparison of Performance of Five Tasks in Bento Box and Oria .....	66
Figure 23 Participants’ Preference Between Bento Box and Oria .....	71

## List of Tables

Table 1 Who are the Test Participants? .....	52
Table 2 Participants’ Search Preference in Bento Box .....	55
Table 3 Task Completed Levels in Bento Box.....	56
Table 4 Participants’ Search Preference in Oria .....	62
Table 5 Task Completed Levels in Oria.....	63
Table 6 Tasks Understanding .....	67
Table 7 Participants’ Preference Between A-Bento Box and B-Oria .....	71

# 1. Introduction

Norwegian academic libraries have implemented the Primo web-scale discovery interface since 2013 (Dabin, E. & Preminger, M., 2018, p.38). Compared to older generation library searching tools (Online Public Access Catalog, or OPAC), Primo is a type of advanced discovery tool in academic libraries, which can be used as easily as using “Google”. With its prominent feature of the single search box on its primary interface, Primo became the most popular discovery tool in academic libraries (Brett, Lierman & Turner, 2016, p.6). Azadbakht’s study showed that most of their users thought Primo was intuitive and they could find adequate results by using the tool (Azadbakht, E. S., 2015, p.23). The configuration and customization of Primo gives it considerable flexibility in its localization.

## 1.1 Background

Oria is the Norwegian version of Primo. In Norway over 80 institutions have employed the Primo discovery interface in their academic libraries (Østeby, S.E., 2018, p.1). According to a few studies about current students’ use of Oria, inexperienced library users were still struggling to use library search efficiently. The University of Oslo Libraries’ intention was to continually assess and customize Oria to improve functionality and user experience. An increasing number of academic libraries have adopted a Bento box interface to optimize their library service. The Bento box search interface provides results from different sources in visually separate boxes on the results page. Based on the possibilities of applying of Bento box search interface, Cole Hudson and Graham Hukill drew a dream picture of future library search interface: “*One page, one search box, results from as many library-resource types as possible*” (Varnum, K., 2016, p.146).

There are some studies on library user search behavior in single search box environments. They typically compared users’ search experience before and after the implementation of discovery tools. Unfortunately, very little research has been published on library user search behavior in Bento box environments since the Bento box style interface appeared in 2005 (Daul, M., 2014; Beswick, K., 2017). The Bento box interface is regarded as the best solution to improve library search, Rochkind claimed Bento box is “*the best cost/benefit way to improve article search relatively quickly with relatively large user benefit*” (Rochkind, J., 2012). However, according to Cory Lown, Tito Siettera and Josh Boyer, “*further research is needed to determine whether combining heterogeneous resource results into a single results list helps or hinders library*

*users as they navigate the library resource environment*” (Lown, C., Sierra, T., & Boyer, J., 2013, p.228).

#### 1.1.1 Primo –One of the Most Popular Discovery Tools in Academic Libraries

Most students expect search engines that are easy to use and resemble “Google” or “Google Scholar” (Lown, C., Sierra, T. & Boyer, J., 2013; Daul, M., 2014; Niu, X., Zhang, T. & Chen, H.-l., 2014). Ideally, by entering keywords or a sentence into a single search box, the system will bring back plenty of search results with the most relevant on the first page, preferably within the top five to ten results. To meet such expectation of library’s users, many academic institutions applied discovery tools for library services, also known as “web-scale discovery services” (Breeding, M., 2007) or “Discovery layers”. These full-service discovery tools like Summon (Serials Solutions), EBSCO Discovery Service, Primo Central (Ex Libris) and WorldCat Local (OCLC) are commonly applied in current academic libraries.

Among these discovery tools, Primo is the most popular (Yang, S.Q. & Wagner, K., 2010, p.706). Primo launched the first version by Ex Libris on May 1, 2002 (Breeding, M., 2007, p.28), and Primo version 3 was released in spring 2010 (Vaughan, J., 2011, p.39). Each academic library that implemented Primo has their own customized interface.

In Norway, Oria is a commonly search engine used in Norwegian academic libraries. Over 80 institutions, research centers or academic libraries in Norway share the collections and information through this discovery tool (Østeby, S.E., 2018, p.1; UNIT, 2020).

Compared to older generation library searching tools (Online Public Access Catalog, or OPAC), Primo is a type of advanced discovery tool in academic libraries, which can be used as easily as “Google”. Through this discovery tool, it is possible to retrieve materials across millions of records of local library collections (such as books, articles, databases and digital archives, etc.), hundreds of millions of scholarly e-resources, including books, e-books, journal articles, reviews, legal documents, open-access repositories and contents of library websites. Users can find or maybe get multiple materials, both paper and electronic version, by using a single search. This tool might bring positive user experiences. However, it also frustrates many users with confusing and overwhelming results (Østeby, S.E., 2018). Although Primo has faceted filtering function menus, users without advanced search skills are still struggling to use it properly and



efficiently (Fox, V. B., 2017, p.9). Additionally, academic libraries have an increasing ambition to supply a unified search box to meet the Google generation students' search behavior. Academic library observed their users not only search library's catalog, articles, databases, but also try to search contents usually showed on the library website (Lown, C., Sierra, T. & Boyer, J., 2013, p.227).

In Norway, Oria has been used in academic libraries since 2013 (Dabin, E. & Preminger, M., 2018; Østeby, S.E., 2018). Nevertheless, there are not so many studies of Oria. Østeby conducted a study based on qualitative interview of 20 undergraduate students about their experiences using Oria. She concluded that students were mostly satisfied with Oria and the single search box. They were, however, frustrated with the overwhelming irrelevant search results, and had difficulties in finding the library's own collections (Østeby, S.E., 2018, p.77).

#### 1.1.2 Bento box – One for All?

Bento box is initially used by Japanese for separating different types of food items in different small boxes integrated in one big box. It is also applied as a metaphor on how to solve search results:

*“The bento box search model involves a discovery interface designed to offer a single box for users to enter a query, which is then simultaneously sent to multiple systems. Results are then organized into multiple panels, each representing the items returned from one service. The term bento box was coined by Tito Sierra, using the metaphor of a bento box keeping different types of food items separate to visualize the organization of different types of information separated into cells of a webpage.*

*A typical bento box would present results from the library catalog, articles from the discovery service, content from the website, and resources from a digital collection within separate panels. The bento box discovery interface would interact with the APIs of each of the target services to populate each panel of the results page. Users can then select any panel to see a more complete list of resources according to the content type of interest”* (Breeding, M., 2018, p.8-9).

Cole Hudson and Graham Hukill described the Bento box interface as follows: *“One page, one search box, results from as many library-resource types as possible”* (Varnum, K., 2016, p.146). Duke University Library had an enthusiastic announcement on their website, where depicted Bento box as a unified search results page *“an increasingly popular method to display search results on library websites”* (Daul, M., 2014).

Tito Sierra is one of the contributors to the Bento box search interface in North Carolina State University (NCSU) Libraries, which became the first to implement a Bento box interface in 2005 (Daul, M., 2014; Beswick, K., 2017). The “Bento box” concept, however, did not get much attention before 2010, when they launched a redesigned interface and presented their successful results. Many academic libraries then followed, and implemented bento-style searches, including Stanford University, Columbia University, Dartmouth College, Princeton University, University of Michigan, University of Virginia (Tay, A. & Yikan, F., 2015; Singly, E., 2016). There are more than 38 academic libraries using Bento box style interface as of November 2017 (Mischo, W., Norman, M. & Schlembach, M., 2018, p.303). Over the past few years, there are even more academic libraries using it or on the way to try the Bento box style interface (Jaffy, M., 2020).

A “Bento style” search interface would improve users’ article search (Rochkind, J. ,2012). Rochkind used a very long blog to analyze the strong reasons to apply the Bento box type interface and explained why Columbia University Library should opt the Bento box style display.

According to NCSU’s transaction log analysis, “*NCSU found that 45% of their user clicks on results were in the ‘Articles’ results, vs 35% in the ‘Catalog’ results*” (Rochkind, J. ,2012). Rochkind pointed out libraries did not spend as much time improving article search options as improving catalog search options. Article search is important to library users, but they do not seem satisfied with the current solutions. Rochkind believed a “bento style” search interface could be the best way to deal with contradiction between users’ desire for one simple search for both articles and books and current technology, which makes it difficult to create usable “blended” results. Rochkind suggested the following reasons to apply the Bento box style interface: improve article search in discovery system; integrate catalog search and article search into one single search discovery system, while catalog will not be locked in the commercial vendors.

Kevin Beswick described Bento box search as “*A shortcut to known items or other search tools*” (Beswick, K., 2017). The ambition is to integrate all the information which libraries can provide, including the library’s own catalog, articles from vendors, databases and contents on the library website (like events, spaces, people, services, etc.), into a united simple search interface by a single search box. Beswick introduced the further development opportunity of their Bento style

QuickSearch, which integrated customized configurable statistics modules to collect usage data for searchers.

## 1.2 Motivation for This Study

The discovery environment in academic libraries still has plenty areas for improvement (Breeding, M., 2014, p.38). Lown, Sierra and Boyer recommended, “*more quantitative and qualitative studies of how users search and expect to search the library in a single search environment*” (Lown, C., Sierra, T. & Boyer, J., 2013, p.240). Emily Singly pointed out in her blog that “*We don’t really know yet if ‘bento’ design has solved anything — more user studies are needed, particularly studies measuring bento vs. non-bento usability*” (Singly, E., 2016). The discovery tool with Bento box interface or without Bento box interface have gradually caused attention. Sigrid Østbye suggested that usability studies of Oria would be useful, since many students were interested in talking about problems of Oria’s interface in her interviews (Østbye, S.E., 2018, p.78). Therefore, the question of “*To Bento or not To Bento*” attracted my attention.

Until now there are no published studies about Bento box in Norway that show that any academic libraries in Norway have implemented “Bento-box” style search results display, although they are keen to observe this phenomenon and keep track of its development. The number of institutions that have applied Bento box interface in academic libraries in other countries has kept on increasing, especially in USA. The success of implementing Bento box interface made some Norwegian libraries curious about this phenomenon.

The University of Oslo Library considers the Bento box style search results display to have the potential to provide users with an improved and intuitive search experience. I have always had a passion for improving the development of search interfaces of academic libraries. Therefore, when I read the suggestion about Bento box in Oria from the University of Oslo Library, I contacted the librarians and discussed with them in September 2019. They gave me much inspiration and good suggestions, which encouraged me to explore the possibility of using Bento box interfaces in academic library catalogues.

The purpose of this study was not to simply evaluate whether Norwegian academic libraries should choose Oria with a Bento box interface or not, but to identify the problems and

opportunities with Bento box. Through a prototype test including observation and interviews, this study will explore the strengths and weaknesses of Oria with a Bento box interface.

This study aims to investigate whether Oria with a Bento box interface is suitable and feasible in Norwegian academic libraries. By talking about the new discovery experience of using Bento box, O’Connell pointed out that “*Understanding how our users search and which boxes they really need will allow us to make programmatic decisions in the future as to which boxes should be shown, based on what the user is searching for*” (O’Connell, 2019). This study results may provide academic libraries with some input regarding Oria with a Bento box interface and hopefully be an inspiration for better library discovery services.

### **1.3 Research Question and Chapter Overview**

This study will focus on the following research question:

How is the search behavior of undergraduate students affected when searching a discovery system like Oria, if results are displayed in a Bento box style?

This research question resulted in some secondary questions, such as: What are the characteristics of undergraduate students searching behavior when searching in Oria? What are the characteristics of undergraduate students searching behavior when searching in Oria with a Bento box style interface? Which interface do undergraduate students prefer? What are the advantages and disadvantages of Bento box? How do undergraduate students understand faceted search results display?

In this thesis, a prototype usability test was designed for users to complete a set of five tasks using both Bento box and Oria interfaces. The method and results of the testing will be presented, as well as suggestions that could contribute to further study in this area. It will also discuss its effect on information literacy instruction.

The thesis is structured as follows: In chapter 2, theory on usability testing and some topics related to Interactive Information Retrieve (IIR) will be presented. This is followed by a quick overview of previous studies about discovery tools, mainly Primo, and usability tests and students’ information searching behavior. Chapter 3 will present the most applied usability test

methods, and a discussion of suitable test methods for my study. Advantages and disadvantages of each method will be summarized at the end of the chapter. Chapter 4 will display the test process, such as tasks design, choosing software, making Wireframes, additional tools for observation, following-up questions and the pilot test. Chapter 5 provides results and analyses. This chapter will focus on findings, mainly discuss the research question: To Bento box or Not to Bento box? What participants experienced the difference? Which one do they prefer? Do they have good reasons? Chapter 6 is the last part of the thesis. Here I will give the conclusions based on further discussion and make recommendations for future studies.

## 2. Theory and Literature Review

Studies that focus on users' information searching behavior is a research field within interactive information retrieval (IIR). Usability studies typically focus on users' experiences related to using a product or service.

In this chapter, I will give a short introduction to IIR and user studies, mainly a summary of three information retrieval approaches and their focus; information search behavior and information seeking models related to my study. Secondly, the most important concepts of usability testing will be addressed, such as users, goals and the profile of the target users in this study, i.e. undergraduate students. I will explain how these concepts and theories have been connected to my study. Thirdly, I will provide an overview of earlier user studies of Primo implementation and Bento box results display style interfaces.

### 2.1 IIR and Information Search Behavior

This study examines how different interfaces of the discovery tool affect undergraduate students' information searching behavior. It also addresses user preferences. Information search behavior aims to understand how individuals, as well as groups, retrieve and utilize information to meet their needs in different contexts by interacting with a search system. Researchers try to find patterns of the information search process, how the results meet their information needs and how relevance evaluation affect their interaction with the system.

#### 2.1.1 Three Information Retrieval Approaches

Figure 1 (Goker, A. & Davies, J.,2009, p.32) is a summary of three approaches to information retrieval (IR) systems evaluation: the system- driven approach, the user-oriented approach and the cognitive IR approach (the latter is exemplified by the IIR evaluation model). A system-driven approach is often referred to as a 'black- box experiment', because it is mainly about observing input and output, and measuring the outcome. Such studies do not address the users. The user-oriented approach shifted focus from system to user. Researchers who applied a user-oriented approach realized the importance of the user and the actual information need, because they regarded the user as part of the system. However, there is no comparatively valid experimental control, which was been pointed out by researchers on IIR evaluation model. In the cognitive IR approach, exemplified by here by the IIR evaluation model, both the user's real information needs, the interaction between users and system, and experimental control are studied together.

**Table 2.1** Summary of three approaches to IR systems evaluation: the system-driven approach; the user-oriented approach; and the cognitive IR approach, exemplified by the IIR evaluation model

Aspect	System-driven approach	User-oriented approach	Cognitive IR approach
The user	<ul style="list-style-type: none"> <li>Is not present</li> </ul>	<ul style="list-style-type: none"> <li>Is present, and is perceived as part of the system</li> </ul>	<ul style="list-style-type: none"> <li>Is present, and is perceived as part of the system</li> </ul>
Information need	<ul style="list-style-type: none"> <li>Pre-designed query = information need</li> <li>Static</li> </ul>	<ul style="list-style-type: none"> <li>The user's genuine information need</li> <li>Potentially dynamic</li> </ul>	<ul style="list-style-type: none"> <li>Simulated information needs + the user's genuine information need</li> <li>Potentially dynamic</li> </ul>
Relevance	<ul style="list-style-type: none"> <li>Objective</li> <li>Static</li> <li>Binary</li> </ul>	<ul style="list-style-type: none"> <li>Subjective</li> <li>Potentially dynamic</li> <li>Non-binary (but often treated in a binary way: recall/precision)</li> </ul>	<ul style="list-style-type: none"> <li>Subjective</li> <li>Potentially dynamic</li> <li>Non-binary (but might also be treated in a binary way: recall/precision)</li> </ul>
Model	<ul style="list-style-type: none"> <li>The Cranfield model</li> <li>Laboratory environment</li> <li>Batch retrieval</li> <li>Experimental control</li> <li>Focus: system performance</li> </ul>	<ul style="list-style-type: none"> <li>No formalized model – depends on research focus</li> <li>Real-life operational environment</li> <li>Allows for successive search sessions, depending on research focus</li> <li>Realism</li> <li>Focus: the user and system use</li> </ul>	<ul style="list-style-type: none"> <li>The IIR evaluation model – allows for different research foci</li> <li>Semi-laboratory/semi-real-life</li> <li>Allows for successive search sessions, depending on research focus</li> <li>Realism and experimental control</li> <li>Focus: the user and the system in interaction</li> </ul>
Performance measures	<ul style="list-style-type: none"> <li>Recall/precision</li> <li>Effectiveness</li> </ul>	<ul style="list-style-type: none"> <li>Recall/precision</li> <li>Effectiveness/efficiency/information searching behaviour</li> </ul>	<ul style="list-style-type: none"> <li>Alternative performance measures + recall/precision</li> <li>Effectiveness/efficiency/information searching behaviour</li> </ul>
Strengths	<ul style="list-style-type: none"> <li>Experimental control</li> <li>Quick and cheap to carry out</li> <li>Applicable for IR techniques under development</li> </ul>	<ul style="list-style-type: none"> <li>Realism</li> <li>Users form part of the test</li> <li>Genuine, potentially dynamic information needs</li> <li>Subjective, non-binary, potentially dynamic relevance</li> <li>Applicable for developed and operational IR systems</li> </ul>	<ul style="list-style-type: none"> <li>Realism</li> <li>Experimental control</li> <li>Users form part of the test</li> <li>Genuine and simulated potentially dynamic information needs</li> <li>Subjective, non-binary, potentially dynamic relevance</li> <li>Applicable for developed and operational IR systems</li> </ul>
Weaknesses	<ul style="list-style-type: none"> <li>Lacks realism</li> <li>No users are involved</li> <li>Artificial information needs</li> <li>Objective, binary and static relevance</li> </ul>	<ul style="list-style-type: none"> <li>No experimental control</li> <li>Resource demanding</li> </ul>	<ul style="list-style-type: none"> <li>Resource demanding</li> </ul>

Figure 1 Three Approaches to Information Retrieval System Evaluation

### 2.1.2 Information Searching Behavior and Information Seeking Behavior Models

According to T.D. Wilson (2000), information seeking behavior is different to information searching behavior. Information seeking behavior aims to satisfy the information need with the purposive seeking for information. In this context, the searcher interacts with information systems. Information searching behavior is the micro-level of behavior employed by the searcher in interacting with information systems. It consists of the interaction of all different levels, from human computer interaction level to the intellectual level. All activities involved into the searching process can be regarded as information search behavior, for example, mouse

clicks, changing search strategy, identifying relevance materials. In my study, mostly I focused on the micro level of information seeking behavior, therefore, I would use the information searching behavior concept.

There are many different information seeking models. However, I will here provide a brief overview of three models which I regarded as most relevant to my study.

Firstly, Broder's standard model of search process. Figure 2 from Broder (2002) illustrates the general information seeking process. This model reflects how a user perform a task, and how a user interacts with the system by the search interface. A task is developed into information need. A user sent the query through verbal form to the search engine, then received results. The user would finally refine the query after evaluating the search results if needed to search again. According to this model, when a user uses a search interface to search information for their tasks, they will first derive their information needs from the tasks, then formulate to query. If they could not find the relevant results to complete the task, they would try to refine the query and put into the search engine again, until they feel success or fail to search.

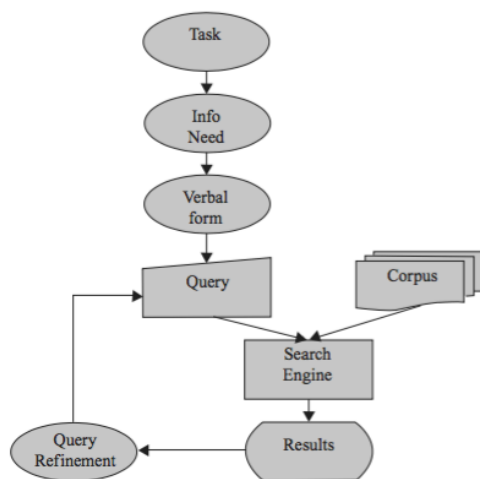


Figure 2 Broder's Model of Search Process

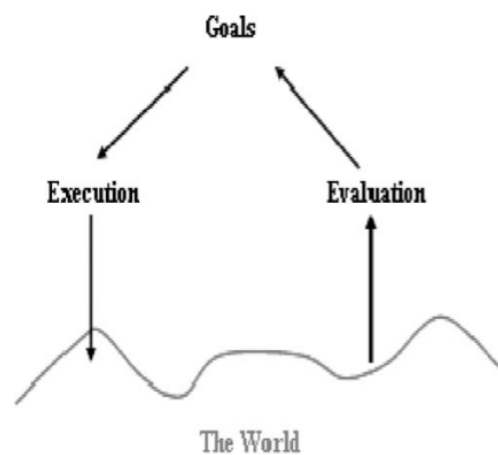


Figure 3 Norman's Gulfs Model

Secondly, Norman's cognitive execution-evaluation model (Figure 3). In this model, Norman divided actions into the doing (execution) and the checking (evaluation) of the result, where he described the gulf of execution as the relation between the usage of the system and the user interface. Queries or navigation in a search system corresponds to executing actions, and



examination of the results to determine whether there is a need to refine the query and search again belongs to evaluation (Hearst, M.A.,2013, p.67).

Thirdly, Bates (1979) grouped four categories tactics to describe a searcher's behavior. Search strategies could make up of sequences of search tactics. The four categories tactics are term tactics, information structure tactics, query reformulation tactics, monitoring tactics (Hearst, M.A.,2013, p.71-72). With these search tactics learning by interacting with the system can trigger new strategies. The interface's learnability can contribute to better interaction between searchers and the system.

## 2.2 Usability Testing

In libraries context, usability testing is widely used by libraries to identify their users' real and potential needs and expectations by studying users' search behavior. Nowadays most libraries have shifted to discovery tools, and most of these products offer a variety of configuration and customization. Usability testing can help libraries to tailor and optimize their search service.

### 2.2.1 Factors to Measure Usability

According to ISO 9241-11:1998 *Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability*, there are three indicators to measure usability: effectiveness, efficiency, and satisfaction. A usability study is not equivalent to a user-oriented evaluation study. Chowdhury pointed out that *“in the context of information science, usability studies involve user studies and the evaluation of information products and services, but user studies are much wide and do not necessarily always aim to measure the usability or performance of a given information system or service”* (Chowdhury, G.G. & Chowdhury, S., 2011, p.14).

User-oriented IR system evaluation studies pay more attention to those factors that can proved to improve service to users while usability studies focus on users' experience of information services. Although *“a usability study is not limited only to the study of the user interface”* (Chowdhury, G.G. & Chowdhury, S., 2011, p.14), Hearst confirmed *“[a]n important quality of a user interface (UI) is its usability, a term which refers to those properties of the interface that determine how easy it is to use”* (Hearst, M. A., 2013, p.5). These points exactly can be reflected in my study. In this study, the research question focuses on what would be the possible effect if Oria with Bento box style interface and functions replaced the existing Oria. This is a usability study while I can take advantage of the user-oriented IR system evaluation approach.

Compared to three indicators to measure usability mentioned in ISO 9241, Nielsen's famous five quality components of usability are more practical and measurable in usability study. In my study, I tried to utility these factors:

- *Learnability: How easy is it for users to accomplish basic tasks the first time they encounter the design?*
- *Efficiency: Once users have learned the design, how quickly can they perform tasks?*
- *Memorability: When users return to the design after a period of not using it, how easily can they reestablish proficiency?*
- *Errors: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?*
- *Satisfaction: How pleasant is it to use the design?* (Nielsen, J., 2003)

### 2.2.2 Users and Goals

Kunnivasky's (2012) described four major steps for a micro-usability test that would be useful as a guideline to plan my usability study. The four major steps for usability testing included: 1) Define target audience (users) and their goals. 2) Create tasks that address those goals. 3) Get the right people 4) Watch them try to perform the tasks. In library context, academic libraries' users included undergraduate students, master students, PhD candidates, academic staff, librarians, and other users. Each library has their own target group, and each group has their own profile. Users' goals will be reflected in the tasks. What's the library's main service? How can users get benefits from that? What do users usually use a product or service to do? Consequently, creating tasks and recruiting participants step depend on the first step. The first step is the most important part for planning a usability testing. In my study, tasks design should reflect the users' information need, their expectation and search behavior, for example, do they usually do known-item search or open-ended search?

### 2.2.3 Profile of Target Audience

As soon as the target audience has been identified, it is important to plan the target audience's profile, which should include demographics (ages, gender, education, income, etc.), behavior, technology user and experience (information competency). It is useful to design the suitable test tasks with the help of profile of target audiences. The profile of target audiences can help define test scenarios, like the usage of cover letter in the concept of simulated work task, which will make tasks closer to the user's natural environment. The profile of target audience can also guide the description of pre- and after-interview guide and observation.

## 2.2.4 User Tasks Setting Scenarios

After identifying of target audience profile and user goals, there is a need to draw a list of usage scenarios, where each scenario is a sequence of tasks performed by users. Based on the results of usability studies of Europeana Digital Library, Chowdhury listed eight different scenarios of user tasks setting:

- *finding texts on a predefined topic*
  - *finding images on a predefined topic*
  - *finding audio and video materials on a predefined topic*
  - *finding materials presenting the same object in different times*
  - *finding materials on a very specific subject (such as a land mark, an event or a person)*
  - *finding materials on a topic of the participant's choice within the context of the general theme*
  - *finding materials on a specific time or event which happened on a specific date*
  - *identifying providers of digital objects who contributed the highest number of objects on a topic, analyzing what was found to be most useful about Europeana and suggesting areas in which material may be lacking*
- (Chowdhury, G.G. & Chowdhury, S., 2011, p.161)

As mentioned above, we can see user tasks included material types (text, image, audio, video), creation date, subject or topic or theme, creators or providers or contributors. In my study, using the library's discovery tool Oria is like using digital library service, the user tasks should include these elements when designing the tasks and building the search process mockup.

## 2.3 Literature Review

Academic libraries commonly conduct user testing on their discovery systems to evaluate potential products or new opportunities for redesigning the interfaces, for example making local customizations when possible, changing products to improve functionality and user experience. User studies on discovery tools involve many aspects, such as users' information needs, users search behavior, implementation of discovery tools for better library services, interface redesign improvement and construction of information competency. In this section, I will give an overview of those studies connected most closely to my study, consequently, the literatures of user studies of discovery tools and Bento Box Search Interface studies.

### 2.3.1 User Studies of Discovery Tools

Kelsey Brett, Ashley Lierman and Cherie Turner pointed out that “*many of the same general usability issues have repeated throughout studies of Primo since 2008*” (Brett, K., Lierman, A. & Turner, C. ,2016, p.20). On the one hand, undergraduate students are the majority group of

academic library user, who are struggling to get benefit of the library's collections; on the other hand, academic libraries continue to follow their own understanding and emphasis to supply library services even as they have realized Google generation undergraduate students' difficult situation. *"Rather than being organized by research needs and user expectations, library information resources have traditionally been organized by the disparate systems library organizations use to manage those resources"* (Lown, C., Sierra, T. & Boyer, J., 2013, p.227). Some usability studies tried to help academic libraries to be aware of this situation (Nelson, D. & Turney, L., 2015; Brett, K., Lierman, A. & Turner, C., 2016). Many questions arise around usability studies about academic libraries: What kind of information literacy instructions should librarians focus on? Can academic libraries only claim vendors' technological barriers prevent academic libraries' development?

Based on the user study of Texas Tech University (TTU) Libraries' implementation of the Primo discovery tool, Perrin, J.M. *et al.* found out that *"most of the users wanted all their search options in one place"* (Perrin, J.M. *et al.*, 2014). Discovery tools resemble commercial search engines like Google, where the users cannot distinguish between the content on the library website, the library's subscriptions and its own collections. Users simply put keywords in a single search box, then expect to get the most relevant results.

Kliwer, G. *et al.* also carried out a usability study of undergraduate students' open-ended searching for a research assignment in the discovery interface Primo. They concluded that undergraduate students had a higher preference for Primo than expected, although students usually encountered frustrations when searching in Primo. This was consistent with previous studies (Nichols *et al.*, 2014) that indicated *"high satisfaction with Primo despite struggling with complex tasks"* (Kliwer, G. *et al.*, 2016, p576). Students were also struggling with using facets to limit the search results.

According to a study of transaction logs (compiled between July 2016 and June 2017) from The University of Illinois at Urbana-Champaign (UIUC) Library shows 56% of the searches were for a known title. Users had difficulties finding the materials that they were searching for, even with the more concrete information in a citation. Mischo, W., Norman, M. & Schlembach, M. observed *"Users perform a large number of copy -and - paste searches of title words, author, and title combined, and full citations—presumably from Google, course management systems, bibliographies, and e-mails"* (Mischo, W., Norman, M. & Schlembach, M., 2018, p. 300). Even

though users often have a material type in mind before they started searching, users usually forgot to use material type to limit the search results after they put in key words and simply got overwhelming results. It typically takes multiple steps to get from a citation to the full-text article or book. However, users took it for granted that it worked as a “Google” search, without understanding the concept of facets to limit results. Thus, the easy simple way to show the search results became more difficult and complex because of unconscious expectations. This showed users’ information search behavior changed by using single search box in library setting.

### 2.3.2 Bento Box Search Interface Studies

Studies of Bento box search interface are relatively rare, mainly divided into two categories: user study and implementation study. The discovery environment still has plenty of areas for improvement, because even known-items search in discovery system works less well than in the catalog. Both students (undergraduate or advanced students) and researchers are frustrated with the overwhelming results, and they are unaccustomed to narrow their search term or filter results. Bento box style results pages, with material types displayed in separate lists, seems more clear and user-friendly (Breeding, M., 2014, p.38). Therefore, many academic libraries have applied “bento-box” style search results display, to improve Primo, especially in USA, UK, Canada and Singapore.

However, there are many different perceptions about this. Some librarians heavily recommended “bento style” search interface, and regarded it as the best strategy to provide significant user benefit with higher feasibility and lower development time (Rochkind, J., 2012). Compared to previous discovery interfaces, Rob O’Connell and Barbara Polowy discussed the new discovery interface (Bento box results display style) at their library by testing based on the first and second year students. The results provided a selection of content that was easier to absorb. They suggested the interface as a “*progressive disclosure of metadata*” (Fox, V.B., 2017, p.9). Dempsey pointed out the advantage of Bento box style results because “*not only may it be difficult or confusing to rank results across different types of resources, but a tabular presentation like this may make more sense to users*” (Dempsey, L.,2014, p.86). In contrast, Shapiro claimed that “*Bento Boxes may have limited efficacy in alleviating user frustration or could potentially make matters worse*” (Shapiro, S., 2018, p.673).

Singley has listed a few reasons for and against Bento box from both sides in the blog post *To Bento or Not to Bento?* (Singley, E., 2016). The reasons for Bento box are: eliminates the

“default search” problem; integrates website search, can help users find any contents on library website; less overwhelming search results display, easily navigate to what users need; makes format types more evident, helping users distinguish format types. The reasons against Bento box are: Bento box may bring small libraries unnecessary complexity; Bento box cannot solve “Below the fold” problem; Bento box doesn’t really remove the “silo” problem, because users still will need to click “more results” for a specific platform if they cannot find the relevant materials on the first display page; until now there are still not enough data to prove its expectation, for example easily to search everything by the single search box. Singley (2016) concluded, *“We don’t really know yet if ‘bento’ design has solved anything — more user studies are needed, particularly studies measuring bento vs. non-bento usability”*.

As the first one who implemented Bento box interface in 2005, NCSU Libraries developed their search interface QuickSearch to a Bento box interface concept (Daul, M., 2014; Beswick, K., 2017). Kevin Beswick described Bento-Box search as *“A shortcut to known items or other search tools”* (Beswick, K., 2017). In 2015, they conducted usability tests on QuickSearch. They concluded that *“[u]sers showed no confusion over the bento-box layout. They seemed to understand it and asked no questions about the layout or arrangement of the search-results page. They all understood they could click through to see more results than the three results shown for each category”* (QuickSearch Usability Test, 2015). They made efforts to develop a unified simple search tool for academic libraries whole content search, towards the phase “full library discovery” which was coined in 2016 by Lorcan Dempsey of OCLC.

An increasing number of universities have decided to adopt Bento box style interface to redesign their website. *“Bento clickthroughs by category indicate a majority of 56.1% for article links..... Only 0.2% of users click on Departmental Library links or librarian contacts”* (Mischo, W., Norman, M. & Schlembach, M., 2018, p.302). Marc Jaffy also shared their Bento-box user experiences at Franklin University recently. To determine whether adopting a Bento-box results display pages would benefit users, they conducted a user study between fall 2018 and spring 2019 at Franklin University. Their conclusion was *“providing bento-box search results would benefit our library’s users”* (Jaffy, M., 2020, p.12).

In Norway, there are few studies about Oria. I only found three papers related to this topic. Two of them are bachelor thesis. One focused on the problems which challenged librarians when Oria was implemented (2015), another one focused on the database implemented (2016) and

the third one is master thesis, which explored students' behavior when using Oria (master thesis, 2018). Mostly they did not pay more attention to the effect of using of Oria's interface. Known-item searching in Primo remains a problematic area. Dabin & Preminger suggested that “[a] new experiment designed to test known-item retrieval could help evaluating this particular aspect and adapting search strategies accordingly” (Dabin, E. & Preminger, M., 2018, p.50).

### 3. Methodology

There are a variety of data collection methods and tools are used for conducting usability tests. But each case has its own methodology. In my study, I must consider several factors, e.g. study objectives, study participants, budget and available time.

Usability studies may be conducted at different stages of the lifecycle of designing a product or services. “*More usability information relating to how far a development of the prototype but before implementation*” (Chowdhury, G.G. & Chowdhury, S., 2011, p.85). Oria has been used in Norway for a few years. Because Primo allows libraries to customize their interfaces to meet end-users’ information needs, librarians in University of Oslo began discussing a Bento box style interface for Oria. Before implementation, a prototype for usability testing needs to be developed. Therefore, my study focuses on differences between Oria and Bento box style interfaces in a usability study of students’ information search behavior.

In this chapter, I will provide a comprehensive review of usability study methods as the starting point for conducting usability studies. Firstly, I will argue usability study methods evaluation, explaining the benefits and limitations of the methods I did not choose, the reason I tried to use card sorting method but finally gave up; secondly, I will present advantages and limits of the chosen methods and explain why I did not use paper prototype as planned at beginning even though I preferred to use it; thirdly, I try to give an outline of methodology in which methods were combined in this study; and lastly, a short description of ethical considerations when conducting usability studies.

#### 3.1 Usability Study Methods Evaluation

According to Chowdhury, G.G. & Chowdhury, S., the usability study tools which were used for conducting usability test in the digital age, include the following methods:

- Card sorting
- Wireframes and paper prototyping
- Expert or heuristic evaluation
- Participatory or user-centered design
- Ethnographic research
- Focus group
- Surveys



- Transaction log analysis (Chowdhury, G.G. & Chowdhury, S., 2011, p.93-97)

### 3.1.1 The Benefits and Limitations of the Methods I did not Choose

This study aims to find out whether users' experience of Bento box display is more efficient and their preference for Oria or Bent box interface. Due to budget, time and feasibility, I discarded some methods. This includes expert or heuristic evaluation, participatory or user-centered design, ethnographic research, focus group, and transaction log analysis. The explanations as follows:

#### a. Expert or Heuristic Evaluation

*“Expert or heuristic evaluation is one of the most cost-effective and useful tools for usability studies”* (Chowdhury, G.G. & Chowdhury, S., 2011, p.95). However, in my study, expert or heuristic evaluation may be not give the answer to my question about usability of applying Bento box display in Oria. An expert who is familiar to the current search interface and information architecture would find it difficult to give a fresh insight on Bento box.

#### b. Focus Group Research

Focus group means to organize a small group of discussions with 6-12 persons led by the researcher or facilitators (Chowdhury, G.G. & Chowdhury, S., 2011, p.46). Focus group as methods of user research have advantages, e.g. gather relevant information quickly, encourage participation and engagement, participants interact with each other, etc. However, focus group methods have some disadvantages, too. It is difficult to get all members of a group participants at the same time and place at the participants' convenience. Besides in a group discussion can make participants' opinions influenced by each other. The result may not reflect their real opinions.

In my study, ideally focus groups should include participants from different disciplines, different levels of students and staffs, it would be quite difficult to gather such a group to do focus groups discussions. As Chowdhury pointed out *“focus groups are good for gathering general qualitative information, but do not produce specific data on tasks, completion time, error rates and so on”* (Chowdhury, G.G. & Chowdhury, S., 2011, p.96). Krug emphasized *“Focus groups are not usability tests”* (Krug, S., 2014, p.112). He explained the difference between focus group and usability tests:

*“In a focus group, a small group of people (usually 5 to 10) sit around a table and talk about things, like their opinions about products, their past experiences with them, or their reactions to new concepts. Focus groups are good for quickly getting a sampling of users’ feelings and opinions about things. Usability tests are about watching one person at a time try to use something (whether it’s a Web site, a prototype, or some sketches of a new design) to do typical tasks so you can detect and fix the things that confuse or frustrate them”* (Krug, S., 2014, p.113).

Therefore, focus group method is not suitable to meet my study objectives.

#### c. Participatory or User-centered Design

*“User-centered design is both a design and a philosophy that puts the user’s needs ahead of anything else”* (Chowdhury, G.G. & Chowdhury, S., 2011, p.95). End users should be involved from the very early stages of the product design and development lifecycle. In my case, the search system is a new generation product which has room for improvement. Different academic libraries have their own collections and end users. It would be difficult to design a universal interface to satisfy all different end users. That is why each academic library make Primo localizations after implementing it. Instead of designing a new product, my study aims to find out whether the alternative interfaces make users be more satisfied with the search system. In addition, each library has their own budget, which limits what they can do to improve it. Therefore, this method is not practical for my case, either.

#### d. Ethnographic research

This method is useful, but time-consuming. It is not practical to develop a complete Bento box system and observe users using it, just to compare two different interface designs. Considering the qualitative approaches used in ethnographic research, such as observation, interviews and tasks, I will integrate them into my main data collection methods.

#### e. Surveys

Surveys are usually conducted by sending questionnaires to users. It can be conducted at any stage of the product lifecycle, but it is better when the product on the market or as a beta test just before official launch (Chowdhury, G.G. & Chowdhury, S., 2011, p.96-97). Most common problems can be detected through analysis of survey data. But more details about the problems, and what was behind the problems they mentioned would be difficult to identify. In my case, the vital problems of Oria without Bento box display are already identified by earlier studies

(Fox, V. B., 2017, p.9). And Oria with Bento box display interface is still a prototype. Therefore, it is unreasonable to depend on surveys.

#### f. Transaction log analysis

Based on the digital footprints of the users, transaction log analysis collects usage statistics, which can be conducted at any time without directly interfering with users. Transaction log has a long history as an analysis tool for library information systems. Therefore, it is regarded as “*an extremely useful and relatively easy method for collecting usability data*” (Chowdhury, G.G. & Chowdhury, S., 2011, p.95). Unfortunately, I am not the employee of academic libraries which implemented Oria, it would be very difficult for me to get access the data and analysis it. In addition, there is only transaction log in Oria, meanwhile, the other interface is a prototype, which cannot have any transaction log. Besides transaction log analysis cannot catch users’ demographical data and supply details about when and why users’ search behavior like that while users’ information needs and cognitive behavior are changeable. It is impossible to measure factors which caused the problems, such emotions satisfaction (Lown, C., Sierra, T. & Boyer, J., 2013, p.231).

#### 3.1.2 Card Sorting or Not?

After careful study and comparing the disadvantage and benefits of each methods, I had chosen card sorting and prototype test, observation, pre- and after-interview as methods. But finally, I gave up card sorting method, in the following I will explain the reasons.

##### a. Why Card Sorting

“*In the earliest stage of the product development lifecycle it is very difficult to conceive the nature of the information product or service to be developed let alone involve users to evaluate it*” (Chowdhury, G.G. & Chowdhury, S., 2011, p.94). Usually card sorting is used in this stage to discover potential users categorize and organize information. “*Card sorting exercises can play a key role in determining the findability of information within an information product or service*” (Chowdhury, G.G. & Chowdhury, S., 2011, p.94).

In my study, we know the users’ information need, i.e. to find and use collections and databases from the academic libraries and their subscriptions. The test is to find out whether Oria with Bento box display is easier to use to meet their information need, which means the collections

and databases are determined, the system (Oria) is determined while the interface can be changed and redesigned to be more plausible to be used. In the other countries, Primo with Bento box display is already ongoing in many academic libraries. But each library has their own interface. My test tries to find out which structure and categorization would be better for Oria.

*“Card sorting can be particularly useful in situations in which you need to come up with a new organizational scheme but don’t know where to start; when you know that the current organizational scheme isn’t working but aren’t sure why; or when you want to test whether one particular organizational scheme is more intuitive than another”* (Spencer, D., 2009, p.2).

Therefore, card sorting would be an efficient and feasible way to get the overview of the users’ favorite categories and organized structure. I had thought to choose card sorting method before I began with prototype design and test.

#### b. Why not Card Sorting?

During the last presentation at master students’ seminar, I got feedback about the card sorting. Some professors and fellow students questioned if it would be necessary to do it and if I have enough time to carry out it. It would be time-consuming. I have spent lots of time to find out if it is practical and necessary to carry out card sort. In the last seminar, I gave the reason as following:

*“Card sorting can be particularly useful in situations in which you need to come up with a new organizational scheme but don’t know where to start; when you know that the current organizational scheme isn’t working but aren’t sure why; or when you want to test whether one particular organizational scheme is more intuitive than another”* (Spencer, D., 2009, p.2).

Therefore, card sorting would be an efficient and feasible way to get the overview of the users’ favorite categories and organized structure. I chose card sorting method before I begin with prototype design and test.

But theory is one thing while what is practical is another thing. The more I read and experiences of the search interface, the more I discovered card sorting would be unnecessary to my research questions. Because when I got access of more than ten university libraries website which applied Primo (or other discovery tool) with Bento box, I was surprised their categories and organized structure have almost no big difference on homepage. This is responded to earlier study results: *“Users demonstrated a great preference for familiarity, gravitating toward terms*

*and resources that were known to them, and even repeating behaviors that had been unproductive earlier in the testing process” (Lierman, A. et al., 2019, p.8).*

Furthermore, I discovered they used the same platform to adapt it. Most academic libraries using Bento box display interface have the same main categories: catalog, books, articles, databases, special collections (this category has different labels at different libraries). For example, a study of 38 libraries with Bento instances shows “*All have Books and Articles sections. Other resources used include website search (17 libraries), Library/Research guides (12 libraries), journal titles (12 libraries), Databases (14 libraries), Digital Collections (11 libraries), Subject Top Picks (10 libraries), and Library Contacts (10 libraries)*” (Mischo, W., Norman, M. & Schlembach, M., 2018, p.304).

Although my project is a usability study, the research question focuses on the differences between two interfaces. The study aims to find out the general benefits and disadvantages of each interface. The choice of categories would impact both interfaces, Bento box or not, in a similar way. Therefore, trying to find better categories and using more understandable labels for each category would be beyond the scope of this study.

### 3.1.3 Wireframes and Paper Prototype

Wireframes and paper prototyping are very common for designers, may not so familiar to everyone who will carry out a usability study. But it is a very useful tool both for usability study and interface design. Wireframes and paper prototypes are paper-based designs, which provide a visual representation of various functions and workflows through simple line drawings. It is cheaper and simpler. Nowadays it is easy to make a mockup of a website by using wireframes with links among various specific functions and features. Chowdhury mentioned “*Wireframes can also be used to conduct usability tests at the very early stage of product design. The value of wireframes or paper prototypes is that some critical usability information can be gathered quickly and inexpensively with minimal resources before the product is produced*” (Chowdhury, G.G. & Chowdhury, S., 2011, p.94).

### 3.1.4 Advantages and Limits of the Chosen Methods

In the study of comparing users’ search behavior of using Oria and Bento box, I will mix

some methods, to create my own methodology for the test. As analysis mentioned above, I will use wireframe online video test, combining observation, pre- and after interview.

Advantages are that such test would be cheaper, quicker and practical. Wireframe or paper prototype has an estranged cousin: participatory design (Snyder, C. 2003, p.14), which means users are designing's focus and will be considered in the whole process. Observation can get the intuitive users' direct experience by watching the mouse clicks on the screen and avoid bias that may occur in other methods of research. Think-aloud method, together pre- and after interview can give more details about participants' information search behavior, and help to discover important problems.

However, due to methodology mentioned above, it would be very difficult for a person to conduct a very strict formal study, which usually a big team (with a few observers, some facilitators, some help to take notes) charged with lots of participants. Therefore, this study can only reflect part of pictures of usability, not the whole one.

### 3.1.5 Do- It- Yourself Testing

Steve Krug suggested a cheaper and a discount testing method – “*Do-It-Yourself Usability Testing*”. He believed testing one user is 100 percent better than testing none, and testing one user early in the project is better than testing 50 near the end. Usability testing is important, but it is not as far away as people believed in the beginning, which must have a usability lab with an observation room behind a one-way mirror and video cameras to record the users' reactions and the screen (Krug, S., 2014, p.115). Furthermore, based on Jakob Nielsen's idea suggested in article *Usability Engineering at a Discount*, Krug encouraged people to do usability testing, because he believed usability testing could achieve the same results with far fewer participants without needing a usability lab.

“*Do-it-yourself tests are a qualitative method whose purpose is to improve what you're building by identifying and fixing usability problems.*” (Krug, S., 2014, p.119). This method expressed the essences of traditional usability test, basically give participants tasks to do, and observe their search behavior, then learn how to improve the usability. At the same time, he confirmed the limitations: the purpose of this kind of testing isn't to prove anything, because proving

things requires quantitative testing, with a large sample size, and lots of data gathering and analysis.

### 3.2 Design Theory and Test Theory in Practice

According to Borlund: “*A simulated work task situation is a short ‘cover story’ that describes an IR requiring situation and provides experimental control by being the same for all the test persons*” (Borlund, P., 2000, p76). The simulated work task situation helps to describe the source of the information need, the context of the task and the problem that the test person should solve. These tasks would make the test person better understand the objective of the search. Borlund explained the function of simulated work task, “*When presented at the start of an experiment the work task situation is supposed to be perceived as realistically as possible by the test persons. This condition entails the sharing of the universe in a cognitive sense by scenarios, test collections and test persons*” (Borlund, P., 2000, p.77).

#### 3.2.1 Simulated Work Task Situation and Tasks Setting

The system-driven and the cognitive user-centered approaches have been criticized because of their bias in experimental setting and study focus. Over the later years, the importance of combining these two approaches has received increased attention. Borlund presents a set of basic components for better evaluation of interactive information retrieval (IIR) systems. These constitutes the experimental setting, and aims to facilitate evaluation of IIR systems in a way that as closely as possible resembles realistic IR processes (Borlund, P., 2000, p.71). Ideally, evaluation of IR systems should solve the conflicts between the system-driven and cognitive user-centered approaches, i.e. by considering issues of experimental control and realism. According to Beaulieu, Robertson and Rasmussen: “*there is a need to simulate a realistic interactive searching task within a laboratory environment*” (Beaulieu, M., Robertson, S. & Rasmussen, E., 1996, p.85).

Borlund proposed that the experimental setting for interactive seeking and retrieval processes should consist of three basic components: “*(1) the involvement of potential users as test persons; (2) the application of dynamic and individual information needs (real and simulated information needs); and (3) the employment of multidimensional and dynamic relevance judgements*” (Borlund, P., 2000, p.72). The test persons should be real users or potential users, who can develop individual and subjective information needs according to simulated work tasks.

Borlund also concluded that there was no significant difference between the real and simulated information needs. It is, however, important to tailor the simulated work task situations to the test persons as closely as possible to the situation that the test persons will engage themselves in. Consequently, one could either substitute real information needs with simulated information needs through the application of simulated work task situations or mix simulated and real information needs. Real information needs act as the baseline to compare to the simulated information needs, which reflect a more general level. By observation of test persons' information search behavior, simulated information needs showed the systems' effect on real information needs. According to Borlund *"the simulated work task situation provides experimental control by being identical for all the test persons, thus enabling comparisons of the test results to be made across systems and/or system components as well as across the group of test persons"* (Borlund, P., 2000, p.72).

Academic library's real users include undergraduate students, master students, PhD students, researchers, etc. The main users group is undergraduate students. Therefore, in my study, I will choose undergraduate students as test participants, who either is real user or potential user. If I try to make the test closer to experimental setting, I should use tasks more related to real information needs or simulated information needs. Most discovery tools offer the possibilities of configuration and customization to help libraries optimize their information service, to meet their own target audiences' needs.

Earlier studies of usability testing and information seeking experiences of academic libraries showed the pattern of students' information needs, information search behavior, and their preference of using libraries' home page. Students preferred to search in the Primo's default settings which is usually via the search box on the libraries' home page, *"at least one-third of the participants did not move beyond Primo's default settings"* (Kilewer, G. et.al. 2016, p.571) and they *"were quick to abandon a search if they did not find pertinent articles within the first page or two of results"* (Kilewer, G. et.al. 2016, p.571). Students have a heavy reliance on default settings (Rose-Wiles & Hofman, 2013, p.153). Students prefer for simple search, especially for known items (Østeby, S.E., 2018; Rose-Wiles & Hofman, 2013). They usually just use titles or author's name as search words. Therefore, it is important to make default settings better to meet their general need, which means the name of items in default settings should be make sense for most student users.



This suggests the display of first page of searching results is very important. In this study, I tried to only make first page of searching results in the prototype since most students will not move to the next pages. However, I put it in the follow-up questions in the after-interview part to follow the possibilities of participants' searching the next pages.

Østbye's study was based on the Norwegian students and Oria. The study shows students' information needs are mostly related to study. To find known-items, printed books, e-books, articles or conference compendia are usually the most information seeking activities via Oria (Østbye, S.E., 2018, p.73). She pointed out that students often used material type, language and publishing date to narrow their research results. They especially disliked book reviews appearing on the first place of the searching results instead of book itself when searching a book. Students expected library's own collections could be appeared on the first place or the first page of the searching results.

Based on these study results, therefore, my test design would take Primo's default settings at first place and adapt it. Tasks design would focus on searching known-items, printed books, e-books, articles and special collections. Faceted functions would be limited to material type, language and publishing date.

Meanwhile, I considered simulated work tasks theory. When I designed the tasks, I supposed most test participants would have similar information needs in their real study life, at least it would not make them unfamiliar to the tasks. Therefore, I choose well-known playwright Henrik Ibsen. I choose *How to Write a Better Thesis*. To find a scholarly journal article, I ask participants to find a latest scholarly article about corona virus.

To avoid too much guidance for test participants, I tried to change the terms to ordinary words in the task descriptions. For example, instead of "peer-reviewed" articles, I changed it to scholarly journal article, since earlier studies show lots of undergraduate students are unfamiliar to the term "peer-reviewed" (Brett, K., Lierman, A & Turner, C., 2016, p.20).

At beginning, I thought about using students' reading lists in the syllabus as tasks. Earlier studies showed the citation format may lead participants to use authors rather than title to do simple search, while "*Simply searching the title in the citation would have returned the item as the first search result*" (Brett, K., Lierman, A & Turner, C., 2016, p.16). Based on this, I gave

up the idea to give citation to test all the known-items search. However, I kept one to see whether students really prefer to use titles or authors as key words to do search. Perrin, J.M. et.al. found in the test “*that users faced the most trouble when searching for articles and opening specific databases*” (Perrin, J.M. et.al., 2014, p.62). Therefore, I tried to use two article search tasks, which of one used citation with details of author, title, name of journal and publishing date, another was an open-ended task with the simulated work task situation.

But how many tasks should be used in a paper prototype test? One team chosen seven tasks (Perrin, J. M., et al., 2014, p59); one group “identified was ensuring users could complete six top tasks in the new site: find a book, find an article, locate a database, place a hold on an item, determine the library’s hours, and locate citation help” (Varnum, K.J., 2016, p163) ; another one used three tasks (Hanrath, S. & Kottman, M. , 2015, p6); even there was one group which only used two tasks (Brett, K. , Lierman, A & Turner, C., 2016, p7 ). But the difficulty of tasks is quite different, some questions are very simple while others are open questions which needed to spend much more time. In addition, their test goals vary from one group to another group.

According to Snyder’s paper prototype test experience, “the time available for tasks is typically the test length minus 10 to 20 minutes” (Snyder, 2003, p136). Considering test participants’ patience and better test results, it should “*reserve another 10 minutes at the end of the test for questionnaires or to discuss interesting issues that arose*” (Snyder, 2003, p136). If “*a task longer than 30 or 40 minutes can be fatiguing for users (not to mention observers)*” (Snyder, 2003, p137), if the tasks are shorter than 5 minutes or less can be not enough for the test.

In my case, I will use Oria’s default and just let participants to do simple search (by using title, author or key words), won’t suggest them to do advanced search. Each participant needs complete two times the same task in two different interfaces.

Especially in current situation, most students sit home alone because of coronavirus pandemic. I don’t know how long students’ patience would be likely to extend for an online test and survey. Therefore, when designing tasks, test designers should consider if the whole test time would be suitable the test situations. Otherwise, tests should be broken down to several test sessions, and divided into several periods. Due to the time, personal and resource limits, in my study, I must limit the tasks and test sessions. Some functions of library website were not included in this

test, for example, checking out the physical collections, getting access and downloading the digital materials, etc.

About magical numbers. Experts on usability testing have argued the magical numbers for formative usability testing, the best-known magic number is 5 (Nielsen, J., 2000). Later discussed they about 8 (Perfetti, C., 2001) or 10 (Hwang & Salvendy, 2010) test persons could make tests have validity. James R. Lewis has a summary about lessons learned of usability. In the article, *Usability: Lessons Learned ... and Yet to Be Learned*, he analyzed the different ideas in the past decades on magical number using in usability tests, and emphasized “‘magic number’ rules of thumb for sample size requirements for usability tests are optimal only under very specific conditions, practitioners should use the tools that are available to guide sample size estimation rather than relying on ‘magic numbers’” (Lewis, J.R., 2014, p.663).

### 3.2.2 Preference VS. Performance

In this study, my research focus is to find out what is students’ preference, not performance. Therefore, the point is not to see how faster or how easily they and complete the tasks, but to see how they complete searching tasks through two different interfaces. Which one would be most participants’ favorite, and why they prefer one of the interfaces would be the central factors taken into consideration when I design tasks.

This test aims to gain a deep understanding of users’ preferences and searching behaviors, to give support the possibility and necessity redesign and implementation of a more useful, easy-to-use product in academic library.

By adapting System Usability Scale (Guay, S., Rudin, L. & Reynolds, S., 2019), I will use a survey (follow-up questions) to get the evaluation of participants. As I mentioned above, the study focuses on test users’ preferences and searching behaviors. The users’ preference for Oria with Bento box or for Oria current version is the key point. However, a task operation analysis using the prototype: how many mouse clicks, time to find and identify pertinent documents; can reflect users’ performance. Instead of measuring time, tasks are measured in number of steps. Accuracy of performance and completion rate will be based on the number of fluent steps and mistaken steps.

Again, I must make an emphasis on this study mostly focus on users' preference of which interface. Tasks performance is limited to find the relevant document corresponded to the task, e.g. participants need not find the physical place or get access to online resources by logging in. In a short, my focus is not test Primo's usability, which is beyond of this study's scope. If libraries attempted to change their interface to improve usability of Primo, there would be another thorough usability test, which may include all the functions.

### 3.2.3 "Think-aloud" Current VS. "Think-aloud" Retrospective Methodology

"Think-aloud" current methodology is where participants are asked to describe what they are thinking and doing, explaining why they acted like that as they complete usability tasks through the system. At the same time, researchers can observe on the spot and evaluate the motivation behind the behaviors afterwards (Lewis, C.H. 1982, p1; Guay, S., Rudin, L. & Reynolds, S., 2019, p 90; Brett, K., Lierman, A & Turner, C., 2016).

"Think-aloud" retrospective methodology is where participants are requested to recall the searching process and explain their information search behavior after they have completed the tasks. *"This method involves participants first carrying out their tasks silently, after which they verbalize their thoughts in retrospect"* (van den Haak, M. J., & de Jong, M. D. T., 2003, p.285). From this method, we can see whether participants changed their mind after they finished the tasks. The problem is that recall maybe not totally as exact as what they were thinking when they were doing it. Earlier studies show both methods *"resulted in similar numbers and types of problems"* (van den Haak, M. J., & de Jong, M. D. T., 2003, p.287). But *"the participants in the concurrent think-aloud condition performed less successful than the participants who worked silently and verbalized in retrospect"* (van den Haak, M. J., & de Jong, M. D. T., 2003, p.287).

I prepared to do face-to-face paper prototype test with "think-aloud" retrospective methodology at beginning, because I did not want to interrupt test participants all the time which maybe give participants too much guides unconsciously, finally may lead to false test results. When I changed paper prototype test to remote online wireframes test, respectively, I changed to "think-aloud" current methodology. When I observed their search behaviors, I asked them to describe and talk aloud what they were thinking. However, I avoided disrupting participants as much as possible by using a think-aloud form (see **A2 Think Aloud Form**) to note the searching process

while I observed them. This method would be flexible for me to observe quietly and efficiently. Participants would feel more comfortable to do searching without being disturbed. Participants spent most of their time talking-aloud and do search naturally as they usually do search on the library website via a single search box.

#### 3.2.4 Balsamiq Cloud Wireframes and Zoom

An IT designer recommended Balsamiq interface design tool. He has been satisfied with the software for many years. It is easy to learn, easy to use and affordable for students to buy one or two months' service (Subscription plans start at \$9/month for a 2-project Space). Besides, they have a one month free trial deal. Most importantly for me, it gave me the opportunity to carry out the remote usability testing. When coronavirus broke out in Norway, and made the whole country shut down, it became impossible for me to do paper prototype testing as planned. Consequently, I regarded Balsamiq Cloud as a useful tool to save my plan of this project.

Balsamiq Cloud is a web-based user interface design tool for creating wireframes, which sometimes is called mockups or prototypes. By using this design tool, designers can generate digital sketches of the idea or the concept for an application or website without coding. The completed wireframes can be used for user testing, getting feedback from stakeholders, or getting approval to start development (Balsamiq, 2020).

Basically, I took screenshots of the Princeton Library (Bento box interface) and the University of Oslo Library (Oria interface) to make Wireframes. Then I connected screenshots together by using link function in Balsamiq, and made the search processes for each task. After free trial, I bought two months' subscription of Balsamiq Cloud service for the test. Zoom is a popular tool to provide users an online meeting. When coronavirus broke out in Norway, most universities here bought the Zoom services for the students and staffs. This makes it easy for me to carry out remote online test, because I don't have to spend lots of time to teach them how to share screen with me.

### 3.3 Ethical Considerations

According to requirement of NSD (Norwegian Centre for Research Data), if the project is going to process personal data, one should notify NSD about the project. On their website, they listed the decisive factors will determine whether the project should be notified or not. I used checklists to go through the possible data I would use in analysis, I did not find any information

which I was going to collect and analysis would be involved personal information. I will not collect personal information, and I am going to process data anonymous. In addition, this meet their requirement of dealing with anonymous information in research project. They explained anonymous information like this: *“If you will only be processing information which is anonymous, it is not necessary to notify us about the project. Anonymous data contain no information that may identify an individual person, neither directly nor indirectly, including a list of names/scrambling key”* (NSD, 2020).

Even it is said it is unnecessary to notify NSD the project if one will only process anonymous information, I felt uncertainty. Although there were no participants who were vulnerable, maybe it was better to notify NSD. Due to the outbreak of coronavirus, and the process of permission from NSD would take time. Since I did not notify NSD my project, I tried not to record the video or voice when later I decided to do remote test via Zoom video meeting. I did not use participant consent forms. In this thesis, I did not use any information that may identify an individual person.

## 4. Test Process

In this chapter, I will present the details of the test process, including tasks design, pre-and after- interviews, making wireframes, preparing video remote observation, recruiting test participants and pilot test.

### 4.1 Tasks Design, Pre- and After- Interviews

The tasks design for this study was derived from several different similar usability studies of discovery tools, specifically studies related to the Primo environment. *Usability Testing for Greater Impact- A Primo Case Study* (Perrin, J. M., et al., 2014, p.58-59) used 7 tasks: “Find an article” was a given topic task, others were known-items, i.e. “Find a book”, “Find an e-book”, “Open a database”, “Find a digital collection item”, “Find a thesis” and “Find an image”. In the article *Integrating Discovery to Improve the User Experience*, Sonya Betz and Ian Robertson used 6 top tasks (top six tasks identified by transaction log), where the first three tasks were known-items, i.e. “Find a book”, “Find an article” and “Locate a database”. Another three tasks were about library information, i.e. “Place a hold on an item”, “Determine the library’s hours”, and “Locate citation help” (Varnum, K.J., 2016, p.163). *Use and Usability of a Discovery Tool in an Academic Library* (Hanrath, S. & Kottman, M., 2015, p.6) used 3 known-items tasks, i.e. “Find a physical book and its location”, “Find an electronic version of the book and get access” and “Search for a given known item: a scholarly journal article and get access full-text”. In *Lessons Learned: A Primo Usability Study*, the comparative test used 2 tasks for each separate test, i.e. test 1 used “Find an article according to a given citation” (known-item), and “Find a peer-reviewed article” (given topic); test 2 used “Find a book and check out from the library” (known-item) and “Find a newspaper article on a given topic” (Brett, K., Lierman, A. & Turner, C., 2016, p.13).

Zhang, T. & Prommann, M. introduced hierarchical task analysis as an expert method to evaluate discovery tools, which was from a task-oriented perspective to complete usability testing. They concluded that “Find a print book”, “Find an e-book”, “Find a journal article” were regarded as three major representative tasks to carry out usability testing of the library websites (Zhang, T. & Prommann, M., 2015, p. 78).

According to the transaction log analysis, Cory Lown, Tito Siettera and Josh Boyer pointed out “thesis” was in the top 20 most frequent queries in fall 2010 (Lown, C., Sierra, T. & Boyer, J.,

2013, p.238), with cumulative frequency was 6.13%. In spring 2011 was “digital+repository” ranked in the 13 most frequent queries with cumulative frequency 4.93%. They concluded that “*The most frequently occurring queries may reveal latent demand for resources and hints about whether the search target meets user expectations. Additionally, ensuring that the most frequent queries provide good results can improve the search experience for large numbers of users*” (Lown, C., Sierra, T. & Boyer, J., 2013, p240). Therefore, I chose to use “Find a thesis” task.

Researchers also noticed that the citation format task seemed to confuse participants, “*they tended to search for the piece of information that was listed first—the authors—rather than the most unique piece of information—the title. If the first search did not return the correct item in the first few results, the participant would modify their search by searching for a different element of the citation or adding another element of the citation to the initial search until the item they were looking for appeared as one of the first few results*” (Brett, K., Lierman, A. & Turner, C., 2016, p.15). Citation format search is one of most common known-item search. It is especially common in article searches, consequently, it should be investigated more as a test task. Although most citations in academic papers can be retrieved directly by the online link to online resources, there is still some citations that need to be searched and found by the users themselves. It is important for students and researchers who need to find the article from the citation to understand how to use citations to search more efficiently. Therefore, I used one article-search task with citation format.

Based on the experiences from the literature mentioned above, which showed usability studies of web-scale discovery services focused on known-item search, I identified 5 tasks: “Find a print book”, “Find an e-book”, “Find a scholarly journal article” (one was known item and another one was topic-based) and “Find a thesis” (topic-based) as tasks for testing. In this study, I tried to combine known- items and open-ended items:

#### Task 1 Find a Print Book

You are going to write a bachelor thesis, and someone recommended you a book *How to Write a Thesis* by Rowena Murray, Open University Press, c2006. Can you find it?

#### Task 2 Find an E-book

Search in the library’s website and find an e-book: Henrik Ibsen’s *A Doll’s House* (Et dukkehjem: skuespil i tre akter), Gyldendalske boghandels forl. København, 1879.



### Task 3 Find a scholarly journal article

Following is a citation from references list of a book, you think it is interesting, you would search on the library website and find out if it is available online.

Harper, Simon. (2007). Is there design-for-all? *Universal Access in the Information Society*. Vol. 6, No. 1, 2007, pp.111 -113

### Task 4 Find a scholarly journal article

Recently almost everyone is talking about coronavirus, your friend asked you to find a latest scholarly article about corona virus in English.

### Task 5 Find a Thesis

This library has a special collection of master thesis about academic libraries. Please find a master thesis about it.

I learned from Perrin, J.M. et al. about the test order of tasks. They pointed out that the order of the tasks was important to test the learnability of the system and minimize user frustration. Meanwhile, learnability of the system is one of important factor to measure usability. The difficulty of the tasks should be increased gradually, “*The method of ‘start easy, finish hard’ is recommended in the CUEP workshop to test memorability and learnability of the website*” (Perrin, J.M. et al., 2014, p.58).

In this study, the first three tasks were known-items searches. The last two were open-ended items, but they were still limited to a few conditions, for example, key words limited to topic, “corona virus” “academic library”. Other relevance evaluation, such as “latest”, “English”, “articles”, “master thesis” were also relatively easy to understand.

Pre- interview focused on demographical questions. I used a pre-interview example described in the article *Testing for Transition* (Lierman, A., et al., 2019, P.90) as a model and adjusted it to apply to my test. The demographic questions included gender, age, education levels, frequency of visiting physical library and using library website. In addition, I added the questions about if they understood “catalog” and “peer reviewed” terms. For the whole pre-interview guide, see Appendix: ***A4 Pre-interview Guide***.

After-interview aimed to measure usability, especially users' preferences, including seven open questions and two closed questions. The after-interview guide helped me to prompt the participants to express their thoughts related to the tasks. In this way, I could track participants' search behaviors while listening to the explanations of their experiences. See Appendix: *A5 After-interview Guide*.

## 4.2 Making Wireframes, Video Observation and Recruiting Participants

Undergraduate students' search behavior was taken into consideration for making wireframes in this study. Firstly, students like to use search defaults (Kliwer, G. et al., 2016; Rose-Wiles & Hofman, 2013), struggle with facets, because of unfamiliarity with facets menu, especially resource types. This study tried to investigate some of the mostly used facets, e.g. resource types, creation date and language. Secondly, students assumed Primo would provide a totally "Google-like" search, but did not realize that the overwhelming search results may need to be refined. They conducted nearly "*all their searches using keywords or search phrases comprised of one to ten terms, whether searching in Primo or not*" (Kliwer, G. et al., 2016, p.570). And they "*were quick to abandon a search if they did not find pertinent articles within the first page or two of results*" (Kliwer, G. et al., 2016, p.571). Therefore, I tried to use search defaults, together with "author", "title" and "key words" suggestions in the single search box for both search interfaces.

### 4.2.1 Making Wireframes

Each task was based on the natural possible searching process. Normally, the users would choose title, author or key words to search (see Figure 5, Figure 6 and Figure 7). The wireframes mocked up the different processes (Figure 8, Figure 9 and Figure 10). When making the wireframe search is limited to choosing "title", "author" or "key words" suggestions to go to the search results pages (as shown in Figure 4, I added the suggestions in the search box). These suggestions exist on the some of Oria search boxes at the home page of library website, for example, The University of Oslo Library. Wireframes listed the different search possibilities. In this way, when returning to the single search box from search results, participants could adjust their search strategies to search again without typing the search words. This allowed participants to have a fluid search experience where the participants could continually change the search strategy, until they found what they were looking for. Although I kept the "Advanced" tab in Oria (Figure 11, Figure12), just as I kept "Catalog" "Articles+" "Databases" (Figure 4) tabs

in Bento box style interfaces, I did not make them link to next layer of pages. In this study, participants were undergraduate students and they probably didn't know how to use advance search.



Figure 4 Single Search Box in Bento Box



Figure 5 Title Search in Bento Box



Figure 6 Author Search in Bento Box



Figure 7 Key Words Search in Bento Box

I tried to remove the logo of libraries to avoid the possible bias. Afterwards I found it unnecessary to remove name of the University of Oslo, because in Norway, almost all the students used the Oria in their institutions and Oria has similar layout. Most participants were not students of the University of Oslo. Therefore, I removed the logo of the Princeton University Library, and left the logo of the University of Oslo in some pages.

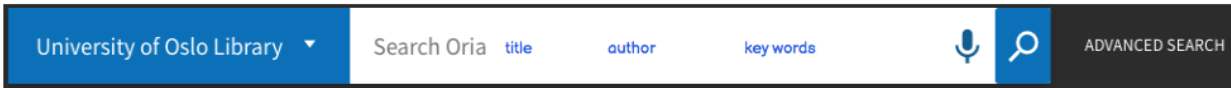


Figure 8 Oria Search box



Figure 9 Title Search in Oria

# Search results for Et dukkehjem

Et dukkehjem

Catalog (20) Articles+ (246) FAQs (2) Art Museum Collections

## Catalog

**Et dukkehjem : forspillet, skuespillet, etterspillet / Einar Østvedt.**

Book  
Østvedt, Einar, 1903-1980  
Skien : Oluf Rasmussens forl., 1976.  
Available Forresta Annex - Temporary » PT3861.03

**Et dukkehjem : skuespil i tre akter / Henrik Ibsen.**

Book  
Ibsen, Henrik, 1828-1906  
Kristiania : Gyldendal, 1921.  
Online [hdl.handle.net](http://hdl.handle.net)  
Available ReCAP » PT3861 .A1 1921

**Et dukkehjem : skuespil i tre akter / af Henrik Ibsen.**

Book  
Ibsen, Henrik, 1828-1906  
København : Gyldendal, 1901.  
Online [hdl.handle.net](http://hdl.handle.net)  
Available ReCAP » PT3861 .A1 1901

**Et dukkehjem. Skuespil i tre akter af Henrik Ibsen.**

Book  
Ibsen, Henrik, 1828-1906  
København, Gyldendal (F. Hegel & søn) 1879.  
Missing Firestone Library » 3352.L.331  
On site access Special Collections - Rare Books » 3352.L.331

**Et dukkehjem ; skuespil i tre akter.**

Book  
Ibsen, Henrik, 1828-1906  
København : Gyldendal, 1910.  
ReCAP » 839.B17 P54

## Articles+

**Et Dukkehjem in Arabic Translation**

Mejdell, Gunvor  
Journal Article - Full-Text Available  
Ibsen Studies  
Vol. 17, No. 1, 2017, pp.28-53

**The Female Jouissance: An Analysis of Ibsen's "Et dukkehjem"**

Anne Marie Rekdal  
Journal Article - Full-Text Available  
Scandinavian Studies  
Vol. 74, No. 2, 2002, pp.149-180

**Three unpublished letters by Henrik Ibsen about the first-performances of 'Et dukkehjem' in Italy**

Lokrantz, Margherita Giordano  
Journal Article - Full-Text Available  
Ibsen Studies  
Vol. 2, No. 1, 2002, pp.59-74

**Pastor Hansen's Confirmation Class: Religion, Freedom, and the Female Body in Et dukkehjem**

Holledge, Julie  
Journal Article - Full-Text Available  
Ibsen Studies  
Vol. 10, No. 1, 2010, pp.3-16

**Pastor hansen's confirmation class: Religion, freedom, and the female body in et dukkehjem**

Holledge, Julie  
Journal Article - Full-Text Available  
Ibsen Studies  
Vol. 10, No. 1, 2010, pp.3-16

[See all Catalog results](#)

[See all Articles+ Results](#)

Figure 10 Title Search Results in Bento box

Henrik Ibsen

Catalog (1457) Articles+ (4420) Library Archives (14) Art Museum Collections

### Catalog

**Henrik Ibsen**

Book

Kultur, Rudolf, 1865-1943  
Leipzig, F. A. Seemann, 1902.  
[Available](#) [ReCAP](#) - [PT8950 .L834 1902](#)

**Henrik Ibsen**

Book

Winterfeld, Achim Adolf Hans Karl Radolf von, 1884-  
Berlin, Friederich, Verlag des Bären-Fischer, 1903.  
[Available](#) [ReCAP](#) - [PT8950 .W558 1903](#)

**Henrik Ibsen**

Visual material

Photographische Gesellschaft  
Berlin : (s.n.)  
[Available](#) [Special Collections - Graphic Arts  
Collection](#) - [D East 53,OC134/Box 03/Wonne-Process  
Printing](#)

**Henrik Ibsen**

Book

Melzer, Oskar F. [Oskar Franz], 1884-1944  
Leipzig, [1902?]  
[ReCAP](#) - [R121 \[Ibsen\] \[Melzer, O. F. Henrik Ibsen\]](#)

**Henrik Ibsen**

Book

Meyer, Hans-Georg  
Völsel : Friedrick, [1967]  
[ReCAP](#) - [PT8950 .M44](#)

[See all Catalog results](#)

### Library Archives

**Henrik Ibsen**

..Henrik Ibsen ... Arthur Symonds Papers. Ibsen, Henrik  
(1828-1905). Series 2: Manuscripts, Typescripts, and ...  
File  
Contained in: Arthur Symonds Papers

**Henrik Ibsen**

### Articles+

**'Seeing ghosts': Gothic contiguities in the  
plays of Henrik Ibsen**

Tiskur, Dipakha  
[Journal Article - Full Text Available](#)  
[Oxford Journals](#)  
Vol. 73, No. 5, 2023, pp.449-457

**Den smilende Ibsen: Henrik Ibsens  
forfatterskab—stykkervis og delt by Ståle  
Dingstad (review)**

Gunn, Oivind  
[Journal Article - Full Text Available](#)  
[Scandinavian Studies](#)  
Vol. 66, No. 4, 2024, pp.473-477

**Hedda Gabler by Henrik Ibsen (review)**

Rain, Laurence  
[Journal Article - Full Text Available](#)  
[Theatre Journal](#)  
Vol. 67, No. 2, 2025, pp.324-325

**Henrik Ibsen auf der Vorstadt Bühne. Carl  
Hempel-Hauswirth's Parodie von «Klein  
Eyoif» (in: Soziale der sozialen  
Raumsemantik Wiens)[Henrik Ibsen on the  
popular stage. Carl Hempel-Hauswirth's  
parody of «Klein Eyoif» in the semantic  
context of Vienna's social space]**

Cornelius Mitterer  
[Journal Article - Full Text Available](#)  
[Studia Austriaca](#)  
Vol. 27, 2023, pp.9-20

**Henrik Ibsens Skrifter, Brev 1844-1905  
[Henrik Ibsen's Writings, Letters 1844-  
1905]**

Hareide, Jonann  
[Journal Article - Full Text Available](#)  
[Ibsen Studies](#)  
Vol. 12, No. 2, 2022, pp.139-152

[See all Articles+ Results](#)

Figure 11 Author Search Results in Bento box

## Search results for henrik ibsen et dukkehjem

henrik ibsen et dukkehjem

Catalog (20) Articles+ (229) Art Museum Collections

### Catalog

**Et dukkehjem : skuespil i tre akter / Henrik Ibsen.**

Book  
Ibsen, Henrik, 1828-1906  
Kristiania : Gyldendal, 1921.  
Online [hdl.handle.net](http://hdl.handle.net)  
Available [ReCAP + PT8861 .A1 1921](#)

**Et dukkehjem : skuespil i tre akter / af Henrik Ibsen.**

Book  
Ibsen, Henrik, 1828-1906  
København : Gyldendal, 1901.  
Online [hdl.handle.net](http://hdl.handle.net)  
Available [ReCAP + PT8861 .A1 1901](#)

**Et dukkehjem. Skuespil i tre akter af Henrik Ibsen.**

Book  
Ibsen, Henrik, 1828-1906  
København, Gyldendal (F. Høgel & søn) 1879.  
Missing [Firestone Library + 3352.1.331](#)  
On-site access [Special Collections - Rare Books + 3352.1.331](#)

**Et dukkehjem : skuespil i tre akter / af Henrik Ibsen.**

Book  
Ibsen, Henrik, 1828-1906  
København : Gyldendal (F. Høgel & søn), 1880.  
[ReCAP + 839.817 P55](#)

**Et dukkehjem ; skuespil i tre akter.**

Book  
Ibsen, Henrik, 1828-1906  
København : Gyldendal, 1910.  
[ReCAP + 839.817 P54](#)

[See all Catalog results](#)

### Library Archives

No library archives results found. Try searching for another topic.

### Articles+

**Three unpublished letters by Henrik Ibsen about the first-performances of 'Et dukkehjem' in Italy**

Lokrantz, Margherita Giordano  
Journal Article - Full-Text Available  
*Ibsen Studies*  
Vol. 2, No. 3, 2002, pp.59-74

**Mottakelsen av Henrik Ibsens Et dukkehjem i rumensk kultur (slutten av det 19. århundre og begynnelsen av det 20. århundre)**

Tomescu Baciu, Sandra  
Journal Article - Full-Text Available  
*Studia Universitatii Babeş-Bolyai - Philologia*  
Vol. 51, No. 3, 2006, pp.115-119

**Noras forestilling om det vidunderlige. Drømmens betydning i Et dukkehjem (1879)**

Waara, Lisbeth Pettersen  
Journal Article - Full-Text Available  
*WordNet*  
Vol. 7, No. 3, 2003, pp.295

**Maskerade-kvinnelighet-frihet: Perspektiver på Henrik Ibsens Et dukkehjem**

Selboe, Tone  
Journal Article  
*Edsår: Nordisk Tidsskrift for Litteraturforskning/Scandinavian Journal of Literary Research*  
Vol. 1, 1997, pp.88

**Hidden Dramas: Cartographic Revelations in the World of Theatre Studies**

Bollen, Jonathan  
Journal Article - Full-Text Available  
*The Cartographic Journal: Cartographies of Fictional Worlds*  
Vol. 48, No. 4, 2011, pp.226-236

[See all Articles+ Results](#)

Figure 12 Key Words Search Results in Bento Box

University of Oslo Library

Including 10 eBooks (10) and 5 versions of eBooks (5)

Standard  Sort by relevance  Personalize

**1** **ARTICLE**  
**Familiens etikk i Et dukkejem (lys avhengig av tittel)**  
 Aarstad, Sissel  
 Norsk litteraturhistorisk og lingvistisk tidsskrift [44] 1963, nr 2, pp.149-152  
[View online](#)

**2** **BOOK**  
**A Doll's House: The wild duck; The lady from the sea**  
 HORN, OLAV  
 Oslo: Universitetsforlaget, 1981  
 Available at UID: Universitetsbibliotek

**3** **ARTICLE**  
**Uansett hva all pleieomsorg er av: The Images of Children in Ibsen's A Doll's House**  
 KANG, KUN  
 Tidsskrift for Læger, 12 October 2005, Vol.125(1), pp.229-232  
[View online](#) PDF file

**4** **MULTI-MEDIA**  
**A Doll's House**  
 HORN, OLAV (1828-1896) (Director)  
 10 versions of this record exist. [Click to see all editions](#)

**5** **MULTI-MEDIA**  
**A Doll's House (play in three acts)**  
 HORN, OLAV (1828-1896) (Author)  
 10 versions of this record exist. [Click to see all editions](#)

**6** **ARTICLE**  
**Uansett hva all pleieomsorg er av: Translation as intervention: Sambha Mitra's Patsi Khela (A Doll's House)**  
 Karmali, Anamika  
 Nord for Tidskrift for litteraturforskning 31 March 2015, Issue 36  
[View online](#) [Open Access](#)  
[View online](#) PDF file

**7** **ARTICLE**  
**Et dukkejem in Arabic translation**  
 Maysal, Samir  
 Arab studies, 2012, Vol. 11, no. 1, pp.27-32  
[View online](#)

**8** **ARTICLE**  
**Uansett hva all pleieomsorg er av: Et dukkejem in Arabic Translation**  
 Maysal, Samir  
 Arab studies, 2012, Vol. 11, no. 1, pp.27-32  
[View online](#) PDF file

**9** **ARTICLE**  
**Ibsen's 'A Doll's House' (Henrik Ibsen) (Critical Essay)**  
 Koster, Paul  
 Tidsskrift for Læger, 1999, Vol.125(2), pp.84-91  
[View online](#)

**10** **ARTICLE**  
**A Doll's House**  
 Tompkins, JENNIFER  
 Ibsen: Nordic Literature and Culture, 2014, Vol.34, pp.12-14  
[View online](#)

**Tweak my results**  
  
 Sort by: Relevance  
 Show only:  
 - Any collection (19)  
 - TransPeer-reviewed Journals (32)  
 - Full Text Online (1,901)  
 - Physical documents (26)  
 Open Access  
 Resource type:  
 - Article (32)  
 - Book (16)  
 - Article (14)  
[Show More](#)  
 Creation Date:  
 from to  
 1800 2020 Filter  
 Language:  
 - English (19)  
 - Norwegian (15)  
 - Norwegian Bokmål (9)  
[Show More](#)  
 Creator/Contributor:  
 - Ibsen, Henrik, 1828-1896 (1)  
 - Karmali, Anamika (1)  
 - Koster, Paul (1)  
 - Mitra, Sambha, 1968- (1)  
[Show More](#)  
 Library:  
 - Universitetsbibliotek Oslo (19)  
 - Universitetsbibliotek (17)  
 - Universitetsbibliotek (10)  
 - Universitetsbibliotek (10)  
[Show More](#)  
 Subject:  
 - Ibsen, Henrik (19)  
 - Language & Literature (15)  
 - Drama (14)  
[Show More](#)  
 Dewey:  
 820.0 (1)  
 820.82 (19)  
 820.82 (17)  
 820.82 (17)  
[Show More](#)  
 Series:  
 - Tidsskrift for Læger i performance i et dukkejem (1)  
 - Et dukkejem (1)  
 - Et dukkejem (1)  
 - Ibsen, Henrik (1)  
[Show More](#)  
 Journal Title:  
 - Arab studies (1)  
 - Tidsskrift for Læger (1)  
 - Tidsskrift for Læger (1)  
[Show More](#)  
 Bibliography:  
 - International journal (1)  
 - Tidsskrift for Læger (1)  
 - Ibsen (1)  
 - Ibsen (1)  
[Show More](#)  
 Institution:  
 - Universitetsbibliotek i Norge (19)  
 - Universitetsbibliotek (10)  
 - Universitetsbibliotek (10)  
[Show More](#)

1

Figure 13 Title Search Results in Oria





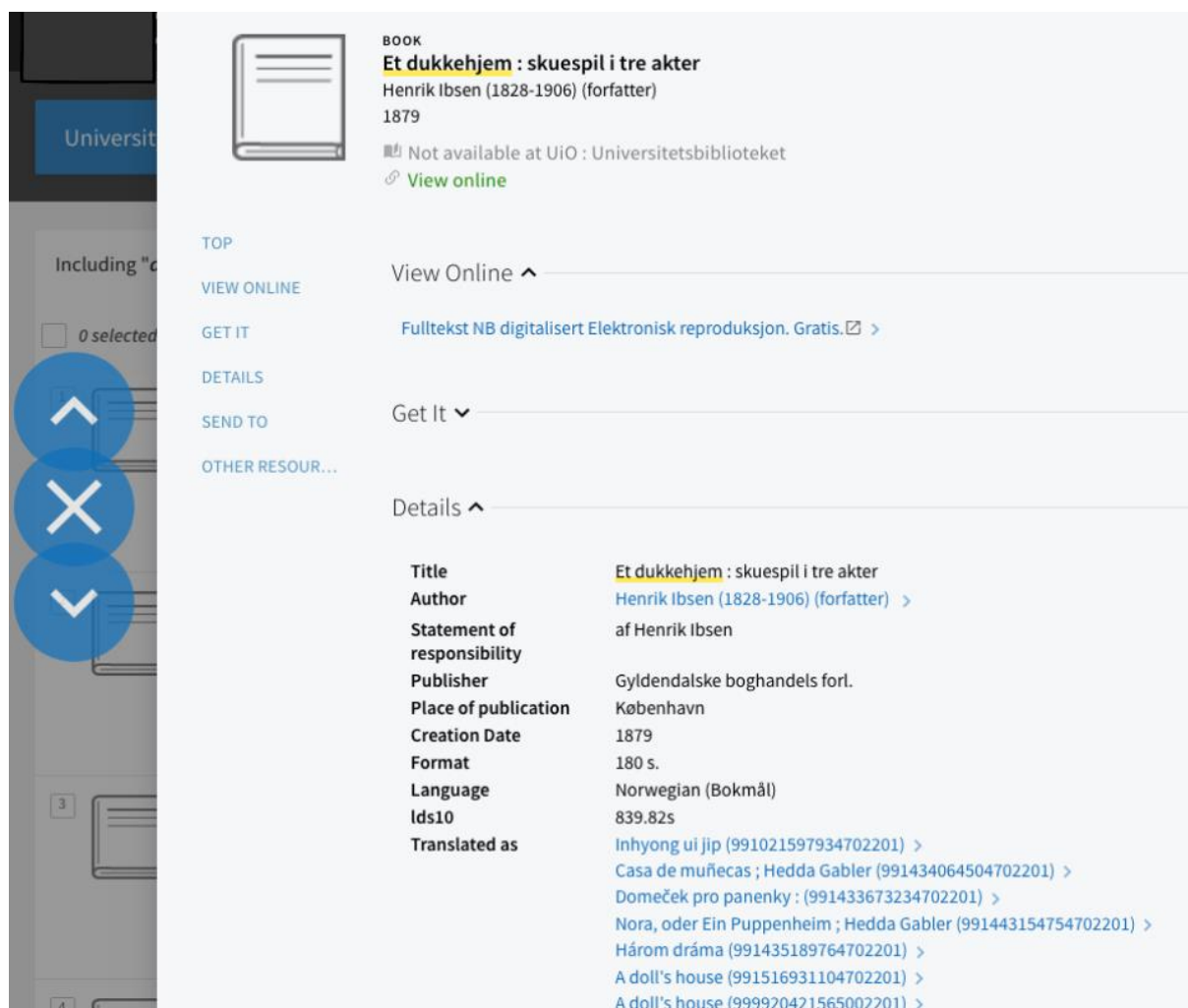


Figure 15 Final Search Result in Oria

Considering the relevance to the test focus, I removed some unnecessary content on the home page, such as contents of library websites, e.g. opening hours, contact librarians, etc. I did not want participants to be overwhelmed by the amount of information on the screen. It may confuse and distract participants from testing. In addition, the first page of search results did not show the contents of library websites in Oria (see Figure13, Figure14). Here I must clarify one thing: I will test a single search box design interface, not the whole library homepage design. However, if Bento box style interfaces are to be implemented with a box supplying the function of library websites' contents results, it will be important to keep all the possible information in the test. Consequently, the task of searching library websites' contents should be included in the test design.

I omitted the links to full-text articles and links to browse journals (Figure 15), because the most important function I aimed to test was whether users could find the resources via the single search box, not the final links or the book shelves. When a user found the item, there is usually a link which could lead to the online full-text or printed materials in the physical library. “One-click to full-text” technology makes it possible to easy get access to the final materials in current Oria. e.g. search results with a “download pdf” icon shown under each article in the search result. Therefore, it is unnecessary to test this function in my study. To get access to the final locations of resources involved in other issues, which are beyond this study. For example, logging in with the user’ account, the library’s subscriptions and the processing of outside databases. Therefore, I did not make wireframes of the links to full-text articles or links to browse journals.

#### 4.2.2 Video Remote Observation

Carolyn Snyder shared the experience of videotaping usability tests and concluded that 90% of the usability test of paper prototype tests videotapes had never been watched. She summarized several reasons to skip the videotaping: the action is slowed down when using paper prototype; the interface changes frequently because many paper prototypes are revised after each usability test; there is less need for analysis because analysis of usability test data always has a point of diminishing returns and the time would be better spent revising the interface or conducting an additional test; it is may be difficult to read the prototype screen because of bad quality pictures of prototype screens. At the meanwhile, she gave some suggestions for using videotape: if there are not enough observers; if observers cannot in the same room, others may need video camera to relay the action; if need for a highlight tape to help finding many people or important stakeholders; if need for a demonstration tape or just in case it would be useful in future (Snyder, C., 2003, p.199-201).

I planned to use paper prototype, it was unnecessary to apply for permission from the Norwegian Centre for Research Data (NSD) because I will not collect personal sensitive information and will use anonymous participants. When I finally changed to online video test, I was not allowed to record the video without NSD’s permission.

Due to my experience as a journalist, I could both observe and take notes efficiently. As Perrin pointed out “*While Morae [usability software] made the observation process easier, the same results could have been achieved through simple screen-capture software, a video camera, or*

*simple note taking by the facilitator*” (Perrin, J.M. et al., 2014, p.57). If I combined prepared observation forms, questionnaires to take notes, it would be practical and reliable for catching the most important problems.

#### 4.2.3 Recruiting Participants

Undergraduate students make up the largest user group of majority users of Norwegian academic libraries. According to the statistic published on the website, University of Oslo has 27 505 registered students and 3000 PhD students, 3804 academic staff in 2019 (UiO, 2020). The Library holdings include 2.925 million books and e-books. Oslo Metropolitan University has 20 635 students in 48 Bachelor’s programs and 38 Master’s programs, 346 PhD candidates (as of 03.05, 2020, from university’s homepage). These numbers can tell us who has the most potential academic library users in the long run. Even the leading research universities in Norway have more undergraduate students than master students, PhD students and academic staff all together. Many usability studies of library websites do not just focus on this group, but recently an increasing number of usability test started to use only undergraduate students. Therefore, in this study, I recruited only undergraduate students.

Recruiting suitable participants is normally time-consuming. In this study, I recruited participants mostly from friends of friends. According to the project plan, I would come to a university library to reserve a test room, where I was supposed to recruit passers-by to participate in the study. Due to the coronavirus outbreak at the end of February, I had to change the schedule and the test method. Firstly, I changed the paper prototype test to online wireframes test; secondly, I changed recruiting method to social media and networking. I posted my recruiting advertisements in four Facebook groups which included 2 faculty groups, one library’s network group, two universities’ network groups, two social help group based on the students. In addition, I posted the recruiting advertisement on my own page of Facebook. Unfortunately, I only got two response from the social media. One group leader offered to help me to ask the students she would have online meeting. I didn’t get any response. Another one was my first test participant from the social media. Then I tried to use my own network. With the help from friends of friends, finally, I got 8 test participants.

Due to the time limit, I could not delay to get more participants. Ideally, I should test with at least ten test participants. Before Easter holiday, I asked one of my friends for help. She is a professor at a university in Oslo. She helped me to send the email to her over 100 undergraduate

students from different faculties. Unfortunately, I did not get any responses from these students. We all understood the situation: do not be disappointed since coronavirus pandemic made almost everyone feels anxious and depressed.

This project's target group included students and researchers at universities. When I decided to narrow the test participants, my advisor strongly recommended limiting the target group to current undergraduate students, because there were not so many usability studies that only focus on undergraduate students, while those students are the majority of the academic library's users. According to this selection criterion, I chose current bachelor study students at Norwegian universities and colleges as test participants. Due to the coronavirus pandemic, all the universities and colleges were totally locked down for almost two months, it was extremely difficult to limit test participants at one university.

There were no incentives for recruiting participants. I had considered using incentives to encourage students to participate. I even thought about using Vipps to send money directly after participants finished testing, which would be a practical way to show my appreciation without worrying about virus spreading. I usually feel payment mechanism in any form may act as an incentive for getting unqualified participants. Such participants may not be appropriate as test person, which may affect the overall results of the study. Therefore, I did not mention incentives because I believed real participants cannot be bought like buying a product.

### **4.3 Pilot Test**

Pilot test started at a week before the real test. Test focused on estimating the real test time and testing video chat, sharing screen technical problems. Several small changes were made to the wireframes in response to breakdowns identified during the pilot test. I was afraid that the screenshots would be not so easy to read because of small fonts showed on the whole screen. It turned out there were no problem for participants to recognize the contents. I discovered some missing links that some participants would try to click. Then I prepared how to guide them to continue the test if that happened, and how to note their search actions. Through the pilot test, I noticed it would be difficult for participants to search on a mobile since the test was designed for a PC screen. I decided to inform them about technical requirements, which included PC with microphone, Zoom, not mobile phone or iPad, when we made an appointment for test.

Testing took place mostly between 30. March to 11. April, the last one happened 28. April. Only the first test used browser-based online video chat. It took long time to connect with each

other. I didn't discover this problem during the pilot test, but decided to use Zoom for the remaining seven tests.

I tried to avoid having the participants prepare for the test, and guaranteed that no one participant got the test links in advance. Firstly, I contacted each participant by e-mail, and made an appointment for testing. Secondly, I printed out think-aloud form and pre- and after-interview guide for each participant and numbered it. Thirdly, I sent Zoom meeting link to the participant on the day we agreed to do test. When the participant appeared in Zoom meeting room, I sent prototype links one by one to the participant via chat dialog or email. When we were waiting for prototype downloading from Balsamiq cloud, I introduced the test and carried out pre-interview. I observed their performance via video chat, took notes, answered any questions they had.

#### **4.4 Observation and Interview via Video Online Test**

I assumed all participants could use and understand the single search box like Google search box. I reminded them "Don't click back to the main menu button in the web browser. You don't need type anything, just move mouse to click what you usually do when searching in the library website". I did not give them any introduction about how to use the search box; only about how to continue the search tasks. I tried to use the links to connect all the tasks together to pursue a more natural searching workflow in the prototype, which means when they finished a task (the item's final page) they would click the item to get to the next task. In this study, final step means the page showed enough described information about the document, e.g. author, title, name of the journal, created time, etc. (see Figure 15) The links to the final downloaded pages ended the task, and linked to the start of the next task. See Figure 15 example of task 2)

How did I define participants have completed a task? How to measure participants' task performance? At the start of test, I asked each participant to indicate which search strategy they usually use to do searching in the library website, such as "title", "author", "key words" etc. If they chose "key words", I would ask them which key words they were going to use to do searching.

I did not time participants' performances with each task, but I noted the whole time in each test. My study mostly focused on the users' preferences, not the users' performance. I let participants evaluate their own performance and preferences. I used the follow-up questions to ask them to

give explanations of their self-evaluations. When I noticed the opposite facts via observation, I asked why they thought Oria was more efficient while they spent longer time to complete tasks in practice. I got a surprise answer: *“Because I am used to it, I feel familiar with it”* *“I felt it was more efficient, maybe actually not, I do not know, it is just my feeling.”* Participants liked the clear and simple (faceted functions) design.

I noted each error or problem of all participants encountered. Each task was labeled as “Completed with ease”, “Completed with difficulty”, “Completed with prompt” or “Failure”.

Based on two wireframes of Oria of the University of Oslo Library and Bento box interface from the Princeton University Library, this usability study was conducted by using eight undergraduate students as test participants. Each participant tried to complete five tasks on both websites (Bento box interface and Oria). Prototype wireframes were made in cloud software Balsamiq, where two websites’ mockups were created using screenshots from both real websites. The logo of the Princeton University Library and part of logo of the University of Oslo Library were covered up. During the test, most participants did not notice the logo of original university’s website.

Based on simulated work tasks’ theory, and combined with earlier similar usability test experiences, I designed five tasks. The tasks included three known- items search, two open-ended search. Among these five tasks, in order to create an experimental test environment as close to a realistic setting as possible, I chose the tasks based on real information needs and simulated information needs, plus potential information needs. The empirical data were collected through Zoom, where the test persons’ search activities were observed and noted, as well as pre- and post-search interviews with the test persons were noted.

## 5. Results and Analysis

The study used a mixed methods approach, mostly based on observation, tasks tracking via think-aloud form, and pre- and after- interviews. Both the verbal narration of think-aloud, observation of the tasks tracking via think-aloud form, together with comments and follow-up questions and deep interview notes provided a rich amount of qualitative data. The following will present the overview of test participants, undergraduates' information search habits by analyzing the collected data, performance of known-item searching and task understanding. Participants' information competency, especially understanding of facets and library terminology and their preference for Oria or Bento box will be highlighted.

### 5.1 Who are Test Participants?

Pre-interview based on a structured interview guide (see **Appendix A4**). There were 10 demographic questions about test participants, which included both closed questions and open questions. Pre-interview was placed after the test introduction. To save the time, I had done this part when we were waiting for the test wireframes to download from the internet cloud before getting ready for testing.

After completing each online video test, I reviewed the notes of what I had written on the think-aloud form, along with additional insights from observation. Firstly, I tried to write down my intuitions and the most important interview results. Secondly, I tried to recall the observation process and add more information about the part I only used a few key words to avoid mistakes. After I has tested eight participants, I reviewed all the think-aloud forms, my interview notes, by coding and grouping to analyze the test results. Responses to each interview question were also coded and ordered with the patterns of behavior and difficulties in completing each task.

I printed out pre- and after- interview guides and think-aloud form to help me taking note when I observed each test participant. Test participants were numbed by interviewed order. Task results and interview completion times were noted, organized into a spreadsheet. Participants' responses to questions were noted and grouped, and success or failure on tasks' completion were tallied.

Because these two tests were based on prototype wireframes, I could not make the two databases were totally same. For example, I didn't make the search results have totally same versions of one book due to two different libraries had different numbers in their collections.

Therefore, it was not necessary to test the efficiency of each website only through one task, or judged by mouse clicks. Based on this, I did not focus on recording how long each test participant spent on each task. Instead, I asked them to describe their feelings after they finished the whole test by searching the same task in two different interfaces.

To my surprise, all the test participants could easily answer the first three questions about comparing two interfaces. Their answers didn't reflect how well they completed the tasks. Some test participants even insisted the one which they were used to was more efficient while they completed the tasks quicker by using the other. This result reflected undergraduates' understanding of the concept of performance and preference.

Among the eight test participants, three are female, five are male. All of them are current students from different Norwegian universities or colleges, both public universities and private colleges. Test participants included two freshmen, three sophomores, two seniors from a variety of disciplines. All of them are full time students. Two of them have minority background and have gotten higher education before, of which one has master degree while another one has another bachelor degree. Except two who both study physics, all others are from different subjects, which included music, history, business, accounting and information technology. The overview of the backgrounds of test participants as the following *Table 1*.

Participants	Gender	Age	Earlier Education	Study type	Study Length	Study Subject
Nr. 1	Male	28	Master	Full time	First year	Information technology
Nr. 2	Female	22	High school	Full time	Second year	Business
Nr. 3	Female	38	Bachelor	Full time	Second year	Accounting
Nr. 4	Male	21	High school	Full time	First year	History
Nr. 5	Female	25	High school	Full time	Third year	Music
Nr. 6	Male	21	High school	Full time	Third year	Physics
Nr. 7	Male	22	High school	Full time	Second year	Physics
Nr.8	Male	21	High school	Full time	First year	Economy

*Table 1 Who are the Test Participants?*



In addition, when asked “Frequency of visiting the university’s physical library”, participants have very varied tendencies, either usually visit very often, or almost never visit the physical library. When asked “The frequency of using library’s website, most test participants said they used very often. It was surprise to find out the two physics undergraduates almost never use library’s website to search for either books or articles related to study. Their alternative is Google. They even mentioned they like to use Google Scholar. When asked how would they do if they could not open the link to the articles through Google or other commercial search engine, they said they would come to librarians’ reference desk for help. Most of them said they would first google the title or author or related items before they came to the library websites. This verified the findings of earlier studies that students prefer to use Google, especially for the freshmen when they did not learn to use library services. It is natural choice for them to search items using Google simple search. As the Google generation, students are used to the simpler, faster Google search. Even when they discovered the benefits of library services, they still used to google the title or author before they came to library website. Discovery tools like Primo still cannot help them discover more if it is needing them to use facets functions.

## **5.2 Information Search Behavior When Using Bento Box**

Because Bento Box prototype test was placed first in the test order, the participants’ search behavior should not be directly influenced by using Oria. However, participants’ search behavior may directly have affected the search behavior in Oria after using Bento box searching the same tasks. Considering that this study aims to investigate how undergraduates’ search behavior is affected by Bento box style research results display, the results suggested participants’ searching behavior when using Bento box would be more valid and trustworthy.

The data collected about searching behavior mostly based on observation of participants performed the tasks and their talk-aloud, note-record via think-aloud form. In addition, the after-interview of participants gave a deep understanding and explanation of their search behaviors.

Task 1 Find a print book, which is a known- item with details about the book. As the *Table 1* showed, three participants used “title” search while four participants used “author”, and only one used “key words”, to begin with the search. This was the easiest task, all participants performed successfully without difficulty, without prompting.

Task 2 Find an e-book (digital printed book), which is also a known-item, both the book's title and author are world-famous. Most participants used "title" search, and only two chosen "key words" search. As *Table 2* shows, all participants performed the task successfully without any problems.

Task 3 Find a scholarly article, which is a known-item with citation form. Unexpectedly, only three participants performed the task successfully while half the participants performed the task with difficulties, one participant even needed prompting. Normally a known-item article search is the easiest task to perform in discovery system. Mostly because they used "author" search for a known-item article. In fact, it is easier to search "title" for an article search in discovery system than using "author" or "title" "author" combination key words. One participant even used "author" with "name of journal" to search the article.

Task 4 and Task 5 were open-ended topic search, and participants used key words to search as expected. But they used different key words combination. Coronavirus is the topic everyone is talking about, but media used different term, therefore, one of participants tried to use "covid-19", some of them used "coronavirus article". Only one participant knew how to search master thesis because he knew their institution has a repository digital collection of master thesis. But in prototype, I didn't make the repository collection link. Therefore, all the participants felt difficult to find the collection. But after self-learning or getting prompt, seven of eight participants could find the thesis collections.

To the question "Did it bother you to have to narrow the search results?", almost all the participants answered "No." without any hesitation. However, they showed the tendency to Google like search by choosing different "key words". If the pertinent materials they were looking for did not show on the first page, even if not within the first a few items, they would quickly give up on the search results without using facets to limit results. Participant Nr.1 almost only checked the top three to five results. Mostly he used title to do search tasks, by prompt, he liked to use creation date to limit. Obviously, they preferred to key words searching. Whatever "title" "author" or "key words", were all equivalent to "search queries" for them. Basically, they typed the search words in the search box, and expected to get the pertinent materials. If the results were overwhelming, which suggested something wrong with search words, then they just changed the different words to try again. participant Nr.2 used to use title to do search, if she could not find the materials for the task by searching title, she would come back to the search box, and try to use title and author as key words to do the search. She didn't

want to go to the next page. Facets were the alien concepts to most participants. I observed most participants did not even consider using creation date to limit the search results.

	Task 1	Task 2	Task 3	Task 4	Task 5
Nr.1	Title	Title	Title	Key words	Key words
Nr.2	Title	Title	Author	Key words	Key words
Nr.3	Title	Title	Title	Key words	Key words /Master thesis
Nr.4	Author	Title	Author--title	Key words	Keywords/institutions' collection
Nr.5	Key words	Key words	Key words	Key words	Key words
Nr.6	Author	Title	Author--Title	Key words	Key words
Nr.7	Author	Key words	Key words	Key words	Key words
Nr.8	Author	Title	Author	Key words	Key words

*Table 2 Participants' Search Preference in Bento Box*

Even though I kept “Catalog”, “Article”, “Database” tabs underneath the single search box, most participants did not notice them. Only one participant noticed the “Article” tab when doing task 3 (article search). In the after-interview, I asked him the reason why he noticed that as one extra follow-up question. He explained that because he had used a library website with Bento box style when he studied abroad. In fact, afterwards I checked the university library he mentioned, which was not Bento box style. It was possible he remembered wrong or the library website changed their interface style.

Task completion is rated on a simple linear scale: 3 points for easily completed, down to 0 points for failure to complete.

Task 1 was easily completed (3 points) by all 8 participants, so it scores 24 points in aggregate.

Task	Number of Participants Completed with Ease (CE=3)	Number of Participants Completed with Difficulty (CD=2)	Number of Participants Completed with Prompt (CP=1)	Number of Participants Failure (F=0)
Task 1	8	0	0	0
Task 2	8	0	0	0
Task 3	3	4	1	0
Task 4	1	3	4	0
Task 5	0	3	4	1

Table 3 Task Completed Levels in Bento Box

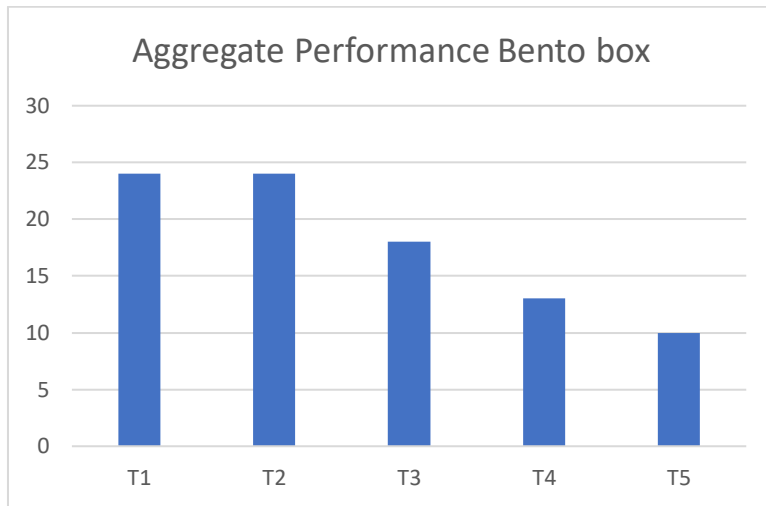


Figure 16 Aggregate Performance of Tasks in Bento Box

University Library **Articles+** academic library  New Search Advanced 0 [Ask Us - Live Chat](#)

142,010,495 results sorted by **relevance**  Include citation-only items

**Content Type**

Filter list

- Archival material (751)
- Audio recording (19)
- Book / eBook (79,641)
- Book chapter (40,552)
- Book review (2,283,762)
- Computer file (1)
- Conference proceeding (76,109)
- Dataset (6,024)
- Dissertation / thesis (708,086)
- Electronic resource (53)
- Government document (7,155)
- Image (25)
- Journal / eJournal (1,958)
- Journal article (9,164,650)
- Magazine (29)
- Magazine Article (5,116,496)
- Map (4)
- Market research (201)
- Music score (1)
- Newsletter (238,957)
- Newspaper (25)
- Newspaper article (161,559,625)
- Painting (53)
- Paper (2,890)
- Personal Narrative (14)
- Photograph (47)
- Poem (12)

1 **online** **Just-in-Time or Just-in-Case? Time, Learning Analytics, and the Academic Library**  
by [Nicholson, Karen P.](#); [Pagowsky, Nicole](#); [Seale, Maura](#)  
Library Trends, 2019, Volume 68, Issue 1  
... for the institution. Aversion to risk is a key feature of bureaucratic organizations such as **academic libraries** (Lynch 1979). Producing the student as future worker, alumnus...  
Journal article: [Full Text Online](#)  
[Preview](#)

---

2 **online** **Assistance Animals in the Library How One Academic Library Developed Best Practices**  
by [Rebecca M. Marral](#); [Barry Trott](#)  
Reference and User Services Quarterly, 09/2016, Volume 56, Issue 1  
... at an **academic library** ...  
Journal article: [Full Text Online](#)  
[Preview](#) [Related Articles](#)

---

3 **online** **Survey of Information Literacy Instructional Practices in U.S. Academic Libraries**  
by [Julien, Heidi](#); [Gross, Melissa](#); [Latham, Don](#)  
College & Research Libraries, 03/2018, Volume 79, Issue 2  
Journal article: [Full Text Online](#)  
[Preview](#) [Related Articles](#)

---

4 **online** **Curation as "Interoperability With the Future": Preserving Scholarly Research Software in Academic Libraries**  
by [Chassanoff, Alexandra](#); [Altman, Micah](#)  
Journal of the Association for Information Science and Technology, 03/2020, Volume 71, Issue 3  
This article considers the problem of preserving research software within the wider realm of digital curation, **academic** research libraries, and the scholarly record...  
Journal article: [Full Text Online](#)  
[Preview](#)

---

5 **online** **Desian effectiveness of academic library web sites : a**

[Back to top](#)

Figure 17 Content Type Facet Option Unobtrusive Button "Apply" in Bento box

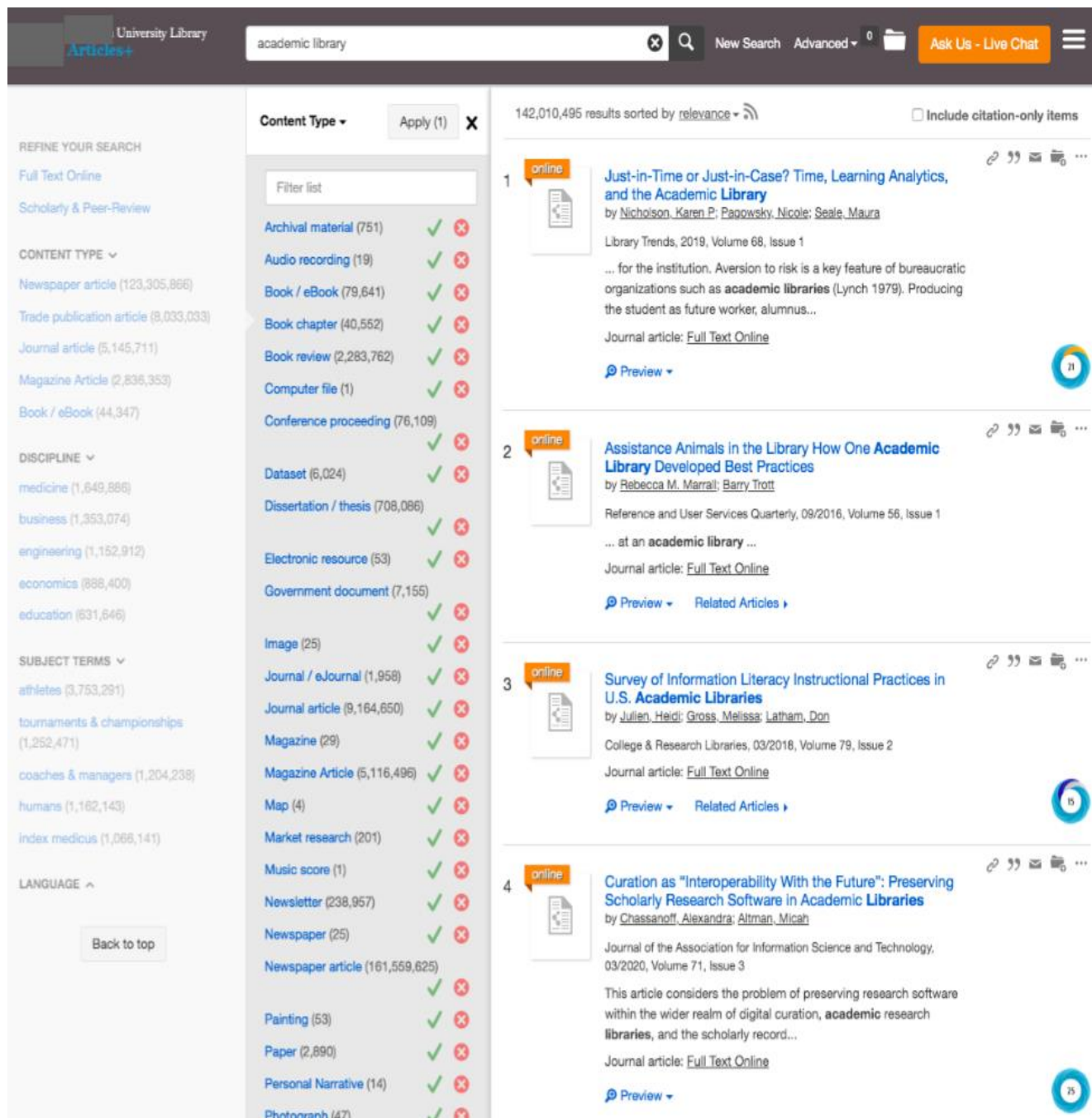


Figure 18 Content Type Facet Option Button "Apply" in Bento box

To sum up, participants liked to use "title" "author" to search for known-items. But there was no big difference of their preference between "title" or "author". For example, participant Nr.2 used "author" to search Task 3 and explained that author was the one she first noticed. Participant Nr.4 chose "author" to search the same task in Bento box because he failed to use "title" to search. He assumed the title could have more search results while "author" search would give narrower search results.

This may suggest they don't know what is the difference between "catalog" search or "article" search. Only one participant tried to click "Article +" tab underneath the single search box. He was supposed to do article search in Bento box. They performed tasks very quickly if they used "title" search. Most participants were not familiar with Bento Box, they didn't pay attention to "Catalog" and "Article" – two big categories. When they could not find the relevant materials, their mouse cursor moved back and forwards, they did not know where to go next: should I go to next page? Where should I click to go to next page? After explanations and prompting, they learned to complete tasks quickly.

They used the single search box like used Google search. They preferred to key words search but they don't know how to choose key words. Most participants trapped themselves in the first search results page. If the pertinent materials they were looking for did not show on the first page, they would easily give up the search results without using facets to limit. After a few attempts, they tried to use facets to narrow the results, however, it was awkward for them to choose which facets to apply, which suggested they need to learn and understand both the system and the terminology of facets and categories.

All participants were excited with the single search box while they also felt frustrated with the facets menu pages. The test was identified one vital usability problem facets as implemented in the Bento box prototype, although it was not about the design of Bento box. All participants failed to click the "Apply" button after they chosen the material type "thesis/dissertation" in the "Article" Bento box section (see Figure 17 and Figure 18). They felt confused with this button because after they chosen the material type "thesis/dissertation", it seemed the results page did not move forwards automatically and bring back the narrowed results. They did not know what was wrong with their action and could not understand the system's design and behavior.

### **5.3 Information Search Behavior When Using Oria**

Oria has been implemented in Norwegian academic libraries since 2013 (Dabin, E. & Preminger, M., 2018; Østeby, S.E., 2018). Most students and researchers are already used to using it. This is the reason I put Oria prototype test after the Bento box style test. The familiarity of Oria could make participants have bias against the Bento box style test. If Oria was tested first, it could reinforce the bias. Oria is a commonplace search engine used in Norwegian academic libraries. Over 80 institutions' research centers and academic libraries in Norway

share their collections and information through this discovery tool (Østeby, S.E., 2018, p.1) Participants were recruited from different institutions, and the after-interview identified that their institutions were still using Oria.

Participants' key words selecting when performing search in Oria could be influenced by searching in Bento box, because they got key words default when they performed the same task in Bento box prototype. But this does not make their searching was more efficient in Oria (*Table 5 Task Completed Levels in Oria*) than in Bento box (*Table 3 Task Completed Levels in Bento Box*).

Even though most participants used the same method to perform the same task in Oria as in Bento box, the failure rate for tasks was still higher than in Bento box. A good illustration of this problem was the response from Task 2. Task 2 showed the typical problems when using Oria's current interface. No matter which search option the participants used ("title" or "author" or "title author" key words search), the first result display page would be mixed and full of overwhelming results. Without narrowing the search results using facets, it was impossible to get similarly relevant results in Oria as the first result page in Bento box. Therefore, it was unbelievable that users claimed to complete Task 2 quicker in Oria than with the Bento box.

Most participants felt familiar with the single search box and the facets menu on the right side of first result page. Unfortunately, they showed confusion with the facets menu and their sub-categories. They could not find the right facets category immediately or directly and took long time to move the mouse back and forth. For example, the button "Sort by" used a dropdown menu including "relevance, date- newest, date- eldest, author, title" (see Figure19, Figure 20). Although it was placed at the top of the facets menu, none of participants were looking for this menu, and it seemed invisible for them. Besides, most participants did not understand material type included more sub-categories, e.g. thesis, dissertation, etc.

Most participants selected "key words" (see *Table 4 Participants' Search Preference in Oria*) to search all the tasks after the test of Bento box. Maybe because the same tasks were used to do both tests, after the doing search test in Bento box, they got fatigued and felt bored searching in Oria, and just wanted to finish the test as soon as possible.

As with searching in Bento box, participants did not know how to choose key words. Three participants tried to use "master thesis" to search in Task 5. One participant tried to use the name of the journal as key words, even though the task gave the citation with the title and the author.



It was interesting to note, however, that after a failed initial search, most participants successfully found the relevant search results. Without being prompted, all participants just searched the default and struggled with the facets functions.

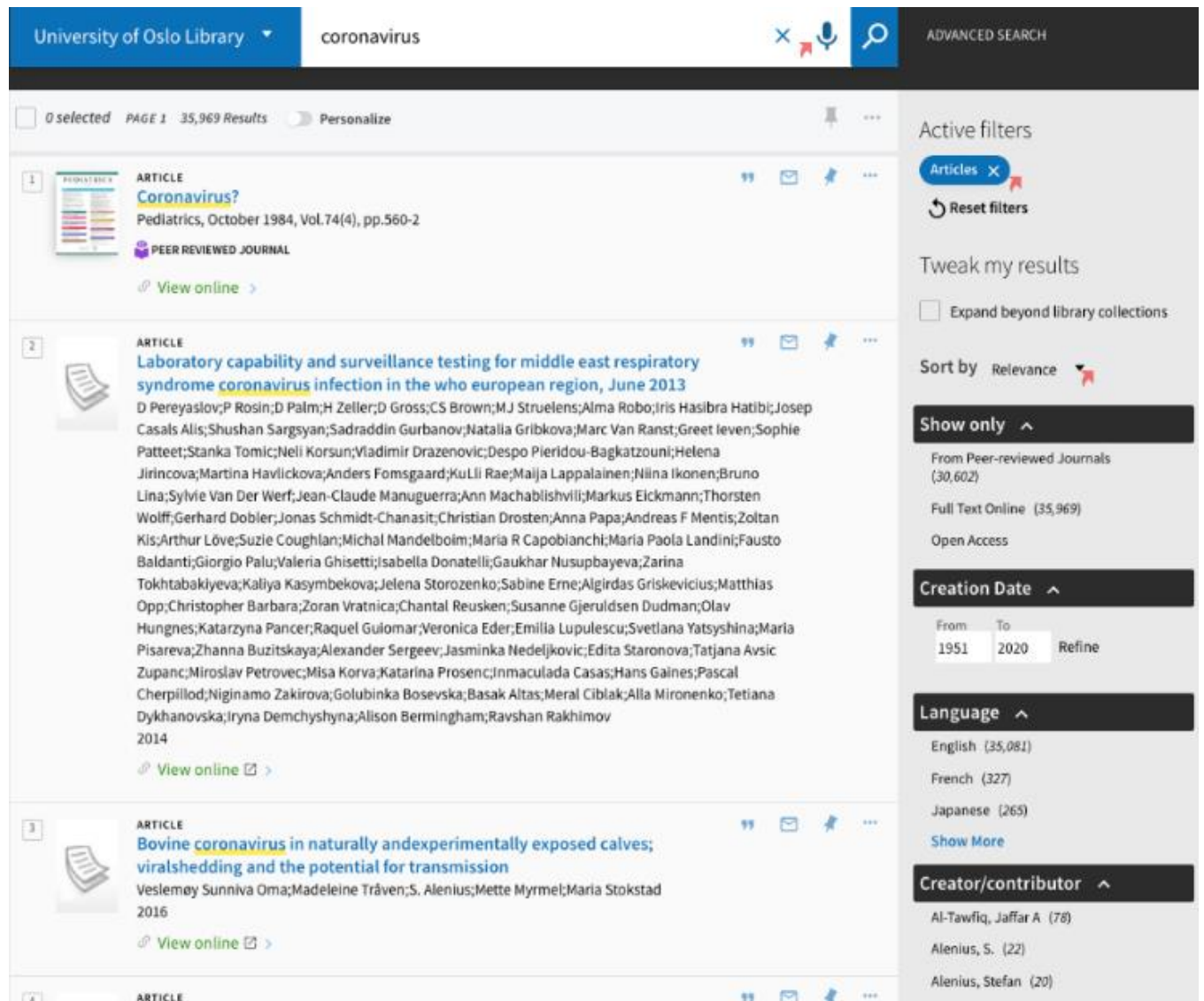


Figure 19 Invisible Button on the Top of Facets Menu in Oria

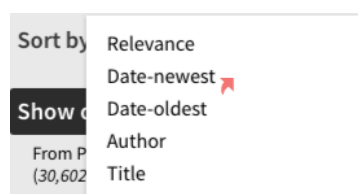


Figure 20 Invisible Drop Down Menu on the Top of Facets Menu in Oria

This finding is supported by other studies. Students often changed their search words after quickly scanning top ten results in the first result page. “*Students were quick to abandon a search if they did not find pertinent articles within the first page or two of results*” (Kliewer, G., 2016, p571). Studies in Primo searches confirmed that “*students completed one open-ended search task and three known-item tasks. The rate of success in the open-ended tasks was higher than for the known-item tasks by nearly 8 percent*” (Kliewer, G., 2016, p.570).

	Task 1	Task 2	Task 3	Task 4	Task 5
Nr. 1	Title	Title-- Key words	Title	Key words	Key words
Nr.2	Title	Title	Author	Key words	Key words
Nr.3	Title	Author-- Title	Title	Key words	Key words*  Master thesis +academic library
Nr.4	Author	Title	Title	Key words	Key words
Nr.5	Title	Author-- Key words	Key words	Key words	Key words
Nr.6	Author	Title	Key words	Key words	Key words
Nr.7	Key words	Key words	Key words	Keywords/English	Key words*  Master thesis +academic library
Nr.8	Author	Title	Articles* Author+ name of the Journal	Key words	Key words*  Master thesis +academic library

Table 4 Participants' Search Preference in Oria

Task	Number of Participants Completed with ease (CE=3)	Number of Participants Completed with difficulty (CD=2)	Number of Participants Completed with prompt (CP=1)	Number of Participants Failure (F=0)
Task 1	6	1	1	0
Task 2	2	1	2	3
Task 3	7	0	1	0
Task 4	0	4	4	0
Task 5	1	0	6	1

Table 5 Task Completed Levels in Oria

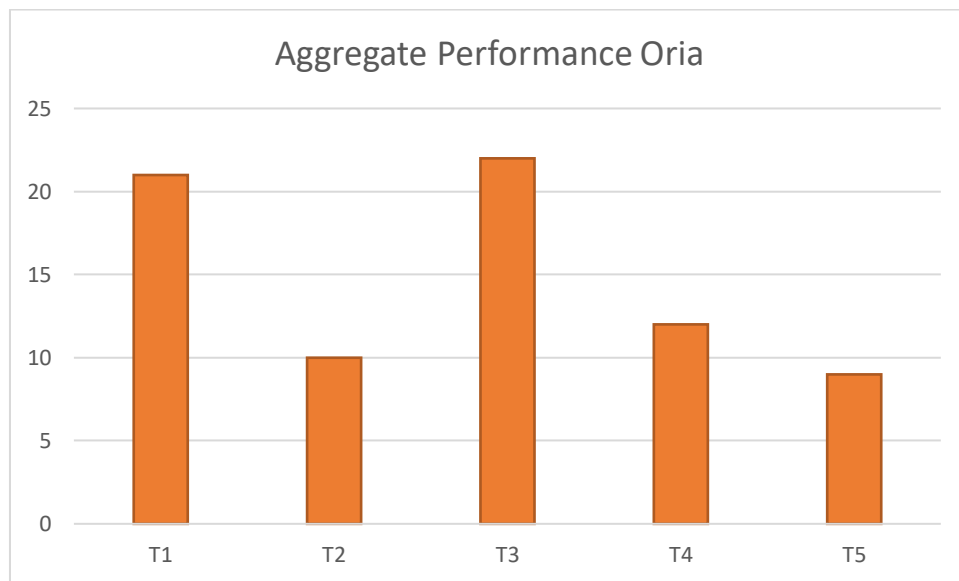


Figure 21 Aggregate Performance of Tasks in Oria

To sum up, this study shows there is no big difference between undergraduate's students search behavior in two different interfaces. Earlier studies of students' information search behavior with discovery tools, especially using one single search box, are consistent with this study and reinforce the study results.

Firstly, earlier studies show most students prefer to simple default search on the library search box (Kliwer, G.*et al.*, 2016, p.579). I didn't make links to the advanced search box. The test results showed participants did have any tendency to use advanced search. I did not observe any participants try to conduct searches in either Oria or Bento box interfaces using the advanced search.

Secondly, participants remained on the first search results page, just skimmed the results page and paid more attention to top five to ten items. They seldom thought about going to the next page. To the question "Did it bother you to have to click to the next page to find the relevant materials?" for most of them, the first answer was "Hold to the first page of results", only one participant said he could come to the second page. After I repeated the question, most of them changed their minds and answered, "Within the second page". Only one participant thought to click to the third or four pages of a search if it was necessary. He explained if there was a need to search about open-ended research assignment, he would click to more pages on the condition that it was difficult to find the relevant materials even after using facets function to limit. He understood it was impossible to find the resources just within the first page. This may have reflected majority participants did not use library search tool very often, without experience of using the facets function.

Thirdly, there was a paradox, where participants would like to narrow the search results while they didn't know how to refine search results. They preferred to use key words search while they didn't know how to choose key words. For example, the first three participants, they were not so familiar with the resource type, therefore, they did not pay attention to resource types, and they don't like to narrow the search result. Instead of facets limiting the search results, they tried to change different key words to do search again and again. It was difficult for them to decide what are the key words to search.

Undergraduate participants couldn't know terms like master thesis, PhD. dissertation, Catalog, peer-reviewed, etc. Therefore, the last and the most important problem they faced is how to understand facets and library's terminology, which academic libraries and vendors have taken it for granted for a long time.

#### **5.4 Performance of Searching and Tasks Understanding**

As mentioned in earlier chapter, I did not focus on measuring performance in this study. This study aims to investigate how undergraduates' search behavior is affected by the Bento box

interface in the discovery system compared to Oria. Search performance would be a useful factor to compare with participants' own experience and their explanation for their search behaviors.

I did not time how efficiently participants completed each task partly due to simple "do-it-yourself-test" design, partly since it was unnecessary for this test. It was difficult for me to observe, time each task performance, while helping with prompt. I tried time each task performance in the last participant test, it was possible, but the time was not so accurate. Even so, I collected the data about how participants' tasks completed levels both in Bento box and in Oria. Based on these, I can roughly compare participants' tasks performance in two interfaces.

As Figure 22 showed, generally four tasks performed better in Bento box than in Oria, only Task 3 performed a little worse than in Oria. Known-items search performed much better in Bento box than in Oria whatever participants used "title" "author" or "key words" search, while participants completed open-ended topic focused search with almost no big differences. If combined Figure 22 with Table 6, an interesting finding appeared: both the most difficult tasks and the easiest task were performed better in Bento box. But Task 3, a known-item article search was performed much better in Oria than in Bento box. If we look back to participants' search behavior when searching Task 3, we notice that participants were mostly using "author" or "key words" search. In truth title search would be the easiest way to get relevant results in Bento box. I noticed this while I searched for all the tasks in Bento box.

Tasks were scored from 0 for "failure", 1 for "completed with prompting", 2 for "completed with difficulty", to 3 for "easily completed". Aggregating points for each task allows us to compare performance of the two interfaces:

The more open tasks (T4 and T5) were more difficult to easily complete in both interfaces (See Figure 22). As Table 6 indicated, three of eight participants regarded Task 2 ("Find an e-book of Henrik Ibsen's *A Doll's House* original version) as the most difficult task. Five of eight participants didn't know how to find a master thesis collection about academic library. Five of eight participants thought Task 1 was the easiest task to complete. Two participants described Task 2 as the easiest. Only one participant chose task 3 as the easiest task, and their judgement was not based on how much easier they could complete task, but based on the task being easy to be understood or they felt familiar to the task. Task 1, task 2 and task3 are all known-item

search. This result showed known-items search was regarded as the easiest task to complete. But through my observation, I still could notice most of them were struggling to find the pertinent materials even it was a known-item, for example, task 2 about finding an e-book of Henrik Ibsen’s *A Doll’s House*. They felt easily to find a known-item with specific details comparatively. To discover more relevant materials for research or study, for example open-ended tasks is even more difficult for them.

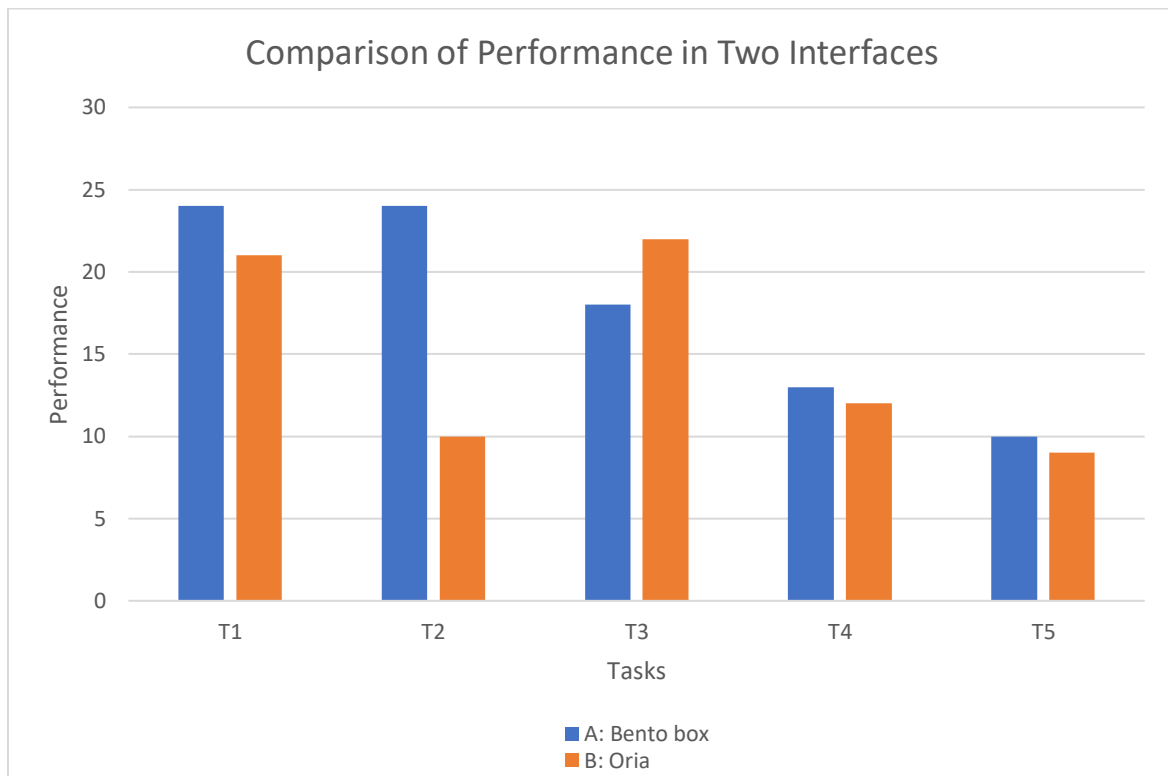


Figure 22 Comparison of Performance of Five Tasks in Bento Box and Oria

This tendency cannot be simply connected to undergraduates’ information competency. Some undergraduates have good information competency, while still feel difficult to search for materials in the library website. For example, one participant studies information technology, and I observed he had difficulties in using facets to narrow the search results, while I noticed he could understand facets very well once I pointed out the facets he was looking for and he began to use the facet function very easily. He concluded that he thought facets were useful to refine the search results even though he didn’t have so much knowledge about facets in library settings.

Participant	Easiest task	Most difficult task
Nr.1	Task2	Task5
Nr.2	Task2	Task5
Nr.3	Task1	Task2
Nr.4	Task1	Task5
Nr.5	Task1	Task5
Nr.6	Task1	Task2
Nr.7	Task1	Task5
Nr.8	Task3	Task2

*Table 6 Tasks Understanding*

## 5.5 Preference for Bento Box or Oria

According to Nielsen’s suggestions about usability test, usability should focus on measuring systems’ learnability, efficiency, memorability, errors and satisfaction (Nielsen, J., 2003). In the after-interview, I used 7 open questions, 2 closed questions to measure usability and users’ preference (see Appendix A5 **After-interview Guide**).

About “Learnability”. *Figure 22 Comparison of Performance of Five Tasks in Bento Box and Oria* anticipated how participants accomplished basic tasks when they encountered the new interface for the first time. It seemed Bento box generally was better than Oria. Task 3 is a known-item search, and participants performed a little better in Oria than in Bento box, which may explain article search function still need to be improved in discovery tool.

About “Efficiency”, Nielsen raised the question “*once users have learned the design, how quickly can they perform tasks?*” to measure system’ efficiency. Oria has been used by all participants at least once while Bento box was quite new. I cannot draw a conclusion just based on such a small sample, but six of eight participants felt Bento box was more efficient. My observation has the same impression although Bento box was tested first while participants spent longer time searching in Oria, on the condition the five tasks they already had completed in Bento box. Even participant Nr.2 and participant Nr.4 felt Oria was more efficient, just because they used to use it and were familiar to the menu. This claim was not based on the

facts but on their feelings. Participant Nr.4 explained he could not remember exactly which one searched faster.

Concerning “Memorability”, participants did not show how easily they could reestablish proficiency after a period of not using in Oria. I heard what they were talking with themselves and observed their embarrassment like *“This website I have used before, but now I cannot find – where it is it?”* participant Nr.4 got a lecture about how to use library and how to search in Oria at the very beginning of the study with the teacher from the faculty. He understood “catalog” “peer-review” immediately after prompting, but he suggested he never heard about them before. He decided how to search very quickly and the mouse moved very fast, but he didn’t use facets function to narrow the searching result at all. Even after prompting, he forgot to narrow the result again. However, both participant Nr.1 and participant Nr.4 liked Oria’s design, specifically they pointed out Oria’s facets menu made more sense for them.

About “Errors”. There were a few severe errors which participants made while searching that were observed both in Bento box and in Oria. There was one severe error observed in Oria: one participant used both “author” and “title” as search words to search Task 2 and she failed to find the item. The error could hardly be recovered without prompting. After the participant tried to use “author” or “title” search, she still could not find the relevance items. For most users, they would assume the library did not have the item or that the system has errors. In Bento box, the “Apply” button (see Figure 17 and Figure 18) for applying facets function in “Article” section would be the severe “error” for users, since no participants realized how to use it.

About “Satisfaction”. Most participants in this study expressed interest in Bento box. They ascribed this to Bento box being more efficient and user-friendly than Oria, even though I observed that they sometime were trapped in the facets menu of “article” or “book” sub-categories. Participant Nr.4 preferred to use Bento box because Bento box felt more user-friendly with less search results displayed in the first result page. Participant Nr.5 explained that Bento box presented search results in a quick and organized manner, and Bento box has a clearer overview of the search results with fewer alternatives. Participant Nr.7 almost never used Oria, and he commented that Bento box was simple and efficient while Oria had too much information and a strange design. Participant Nr.1 explained Oria was familiar and had lots information he could try to narrow the search results. He showed high appreciation for clearly presented side bar labels. He said the most efficient one was the Bento box. But Oria was more user-friendly. Familiarity played the important role in his opinion. His opinion represented



several users' thoughts: Websites should be more stable, the interface does not have to change very often, because the users need time to get used to it.

Evidence showed that Bento box style search result display may be more effective, as users were typically able to do a successful keyword "Google" like search for a task and retrieve articles or books. Having "Article" and "books" sections displayed separately on the first search results page made the results more obvious with section labels. These points may be corresponded to the following suggestions from earlier study:

*"One of the most interesting observations from this study was that most of the users wanted all their search options in one place. They preferred one dropdown menu to handle all their needs. For future development on these kinds of systems, this kind of preference should be kept in mind. The majority of patrons might be happier with a tool with less capability and simpler options rather than a complex tool with many different ways to approach their search"* (Perrin, J.M. et., 2014, p.58).

## **5.6 Undergraduates' Information Competency: Understanding Facets and Library Terminology**

While most tasks were successfully completed by most participants, this study confirmed previous findings that most undergraduates do not understand the complexities of library search and identified several interface issues: (1) they like to use a single unified search box, but are unwilling to use multiple facets for filtering search results. They prefer to type into one or two key words to start a search, if the item did not appear in the first result page, they will immediately abandon and go back to the search box and change key words to try again. (2) lack of understanding of the overview of options to sort search results, and (3) the difficulty in understanding the library's common terminology. Even they tried to make efforts to sort and narrow the overwhelming search results, it took them long time to find it or they did not know where they should move the mouse at all.

Considering about two-thirds of students used a facet in one or more of their searches, study from Kliever, G. et al. suggested *"Overlooking facets that would have improved their search was a common issue"* (Kliever, G. et al., 2016, p.571). Some participants experienced specifically that they found, both in Bento box and in Oria, the interface with facets menu to be overwhelming. Most indicated facets were useful to filter the overwhelming results and helpful to find relevant results. Most participants realized it was impossible for them to complete all

tasks just using single search. They agreed if they were to find the materials in an overwhelming results page, it was important to use facets menu to narrow the results.

Therefore, it is necessary to recommend that librarians focus on helping students develop good search terms, learn library terminology, and use facets to narrow search results effectively.

### **5.7 Findings: To Bento or Not to Bento?**

There are two interesting findings in this study. One is that participants have a strong preference for the Bento box website (Princeton University Library). Participants who preferred the Oria website without Bento box search result display (The University of Oslo Library) explained their reason was their familiarity with Oria. To my surprise, these two findings have coincidentally been reported in another, similar Bento box user experience study at Franklin University (Jaffy, M., 2020) when I ended my testing and was almost finished with my first draft.

Another surprise is that participants did not notice the remarkable difference between Bento box design and Oria, which indicates that participants probably could not distinguish the difference between library's own collection and subscriptions. Most of them did not really understand what "catalog" means. They were struggling with using facets functions which gave them the impression that the two interfaces were similar. There was even a suggestion sentence "*Search for library materials and website content*" in the single search box in the Bento box. No one realized the differing functions of the single search box in Bento box and Oria. The suggestion sentence for explaining the functions became invisible. From this point, earlier studies claimed that "*We need to move from 'full collection discovery to full library discovery'*" (Daul, M.,2014) would be need further discussed. Or maybe because I didn't let them search library contents, for example, opening hours and library events. This type of task should be included in future testing.

*Table 7* shows the summary of the after-interviews of participants, A for Bento box, B for Oria. These results indicated most participants had strong preference Bento box to Oria, and regarded it as efficient and user-friendly. It was interesting that participant Nr.1, Nr.2 and Nr.8 preferred to Oria. Their common reasons were their familiarity to Oria, even Nr.1 and Nr.8 felt Bento box was more efficient. The basic reason was that they felt Oria were familiar to them and they used to use it. At beginning, I guessed such a result may be partly because of the test order. From participant Nr.3 onward, the remaining participants used the same order, all of them were

tested Bento box first, then Oria. They could not see the two links before the test. Among the five remaining participants, only participant Nr.8 agreed with participant Nr.1 and Nr.2, sharing the same opinion: preferred Oria to Bento box, even they experienced Oria was not the efficient one in practice.

Participants	Preference	Efficient	User-friendly
Nr.1	B	A	B
Nr.2	B	B	B
Nr.3	A	A	A
Nr.4	A	B	A
Nr.5	A	A	A
Nr.6	A	A	A
Nr.7	A	A	A
Nr.8	B	A	B
	5 A / 3 B	6 A / 2 B	5 A / 3 B

Table 7 Participants' Preference Between A-Bento Box and B-Oria

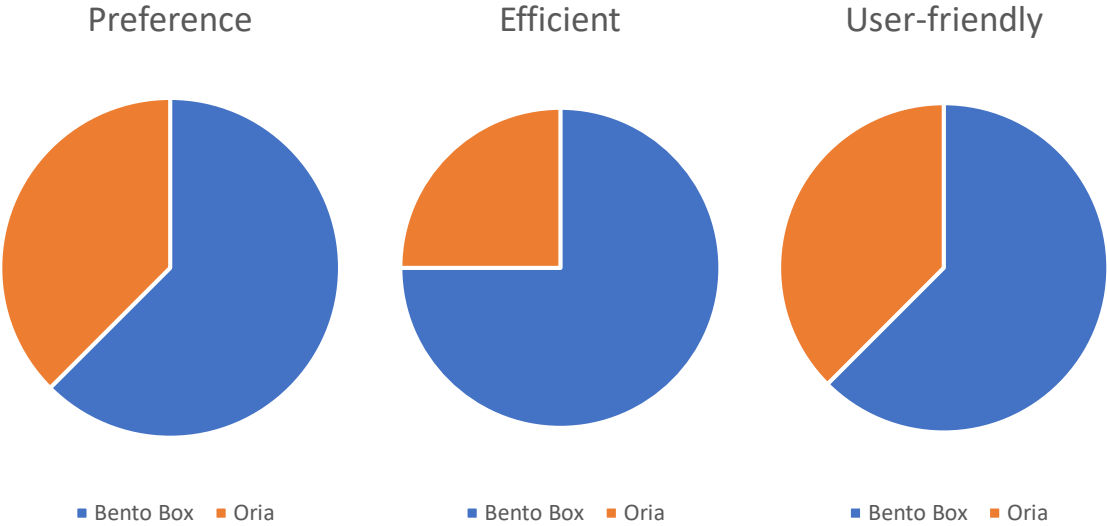


Figure 23 Participants' Preference Between Bento Box and Oria

Participant Nr.4, Nr.5, Nr.6 and Nr.7 didn't use library website very often, some of them even clearly claimed there were no big difference between two interfaces, however, they still preferred Bento box to Oria.

They confirmed that they felt Bento box presented search results in a quick and organized manner. With Bento box, there is a better overview of search results. Participants preferred to use title search to find the relevant materials immediately, and they didn't have to use facets filters to refine the first search results. Participants didn't have to click the next page to check the relevant details. The relevant materials were easily found. In Bento box with title search, participants usually didn't need to the second search sub-categories. They felt more comfortable using Bento box.

Participants were not so familiar with the terms in Oria to limit the overwhelming results. They did have enough knowledge of the facet functions. The primary goal is to give Norwegian academic libraries insights into design and implementation of more user-friendly searching systems. Based on the data collected from the tests as analyzed in previous chapter, I will list the benefits and disadvantages of Bento box and Oria.

#### 5.7.1 Benefits and Disadvantages of "To Bento"

Benefits of Bento box: most participant indicated that Bento box interface was intuitive and gave a simpler, clearer outline, better overview of the layout, less duplicated results. Participant Nr.3 thought it was more efficient, user-friendly, simple and easy to find. Participant Nr.4 appreciated Bento box as a user-friendly interface. Participant Nr.5 noticed that there were more buttons on the first result page to click, would be easy to go to the next search result page. Participant Nr.6 said it was efficient, user-friendly with better relevance rank and better refining menus. Participant Nr.7 praised it as simple, efficient, user-friendly with better layout. Participant Nr.8 simply referred to it as efficient, even though he preferred Oria.

Disadvantages of Bento box: the majority participants did not make any negative comments about the Bento box, and they even did not notice the differences between the two interfaces. Only two participants expressed dissatisfaction with their experience of Bento box. Participant Nr.1 said directly, "*I have a little bad feeling with Bento box. But I still prefer it to Oria.*" "*Bento box has bad facets menu design.*" Participant Nr.2 was strongly against a change to Bento box, because she was familiar to Oria. She disliked the section links in Bento box. She pointed out there were three or four section links to the same next search result page, which confused her and annoyed her. This participant believed there were too many links in each of the tabs, which

would be a challenge to new students, and lead to confusion and unsuccessful searches. Participant Nr.8 was the only one who noticed “Article” tab under the search box in Bento style mockup. When he was doing article search, his mouse hesitated to move back and forth round the search box, I asked him what he was going to do, he said he would click “Article” and got into “Article” sections. This was respond to Rochkind, J. observed earlier, “*In the current environment, many users much of the time just want to be able to ‘search for articles’ without these extra steps*” (Rochkind, J. ,2012).

#### 5.7.2 Benefits and Disadvantages of “Not to Bento”

Facets menu is both a benefit and a disadvantage for Oria. Some participants praised Oria for better design and being more efficient while other participants argued that Oria was too complex with overwhelming results display, and bad relevance ranking. Participant Nr.4 regarded Oria as easier to narrow the search results with precise filter functions. Participant Nr.6 ascribed it to familiarity to Oria and indicated that Oria was “*easy to see (to learn) how to use the functions*”. Participant Nr.1 especially appreciated the facets menu display on the top of the search results pages, and thought it was simple, clear and easy to refine the results.

Disadvantages of Oria: participant Nr.3 blamed Oria as too complex in design, difficult to find an e-book. Nr.4 had familiarity of Oria, and still found it a little difficult to have a result display overview. Participant Nr.6 argued that Oria had a very bad relevance rank, for example, finding *A doll’s house* was so difficult. Participant Nr.7 commented Oria had strange layout, although with more choices to limit the search results, but there were too many to choose between.

#### 5.7.3 Reality and Expectation

Academic library users often used the format facet as keywords, the format was used as a keyword more often than it was chosen as a facets value, e.g. dissertation. Sean Hannan, Steven Heslip and Kristen Johannes concluded in *Fixing Facets*, “*Facet-based searches accounted for 20 percent of total Catalyst catalog searches; 80 percent of searches used a query-only strategy, where users did not select any facets and entered in a query string, typically 4-5 words and longer*” (Varnum, K.J., 2016, p.175). I also observed this phenomenon in my test. Participants stopped midcourse, returned to the search box on the library’s homepage, changed to “author” or “key words” to perform the search again. But even they changed to the combination of different key words search, it was still impossible for them to complete the task without prompting. This would likely address undergraduate students’ confusion with search words and

facets.

The same tasks have been used to search in the second website (Oria) may have prompted students to respond in ways they would consider “correct” rather than working as they did without being prompted. Even as I emphasized the searching process had no right or wrong, it was possible for participants to try to correct their search behavior based on the search experiences in the first one (Bento box).

Both Bento box and Oria interfaces tried to bring users better library service, easy to use with better relevance rank. Both users and libraries dreamed that a single unified search box could solve all the problems which users were struggling with. However, facets and library terminology are unavoidable for users who want to search for materials in practice, even for known-items search. As researchers pointed out “*participants appreciated the pre-search filtering options, and requested more of them (such as an author search), while post-search facets were underutilized*” (Brett, K., Lierman, A. & Turner, C., 2016, p.20). As long as academic libraries still have integrated with multiple platforms (catalog, article, library website, etc.), and users cannot find the relevant materials on the first results page, there will still have facets and next page, next layers’ problem.

## 6. Conclusions, Limitations and Further Study Suggestions

The purpose of this study is to observe how undergraduate students use two different discovery interfaces, and how they experienced these two prototype interfaces, one of which was based on the English version of Norwegian Primo (Oria of the University of Oslo Library), the other was based on the Princeton University Library (Bento box). Each student was observed and interviewed for about 40 minutes via Zoom video meeting with “Share screen”. Eight participants were from a variety of disciplines from different universities or colleges in Norway.

### 6.1 Conclusions

This study indicated that most current undergraduates in Norway had stronger preference for Bento box than expected. Although all the participants had previous experience using Oria, fewer expressed preference for it. On the contrary, five of eight participants said they were in favor of searching in Bento box while only three insisted on their preference for Oria. Considering their familiarity with Oria may bias participants’ preference, the overall result may even more strongly indicate their preference for Bento box. Their preference for Oria did not rely on better performance in searching tasks, but totally depended on their familiarity of Oria. In addition, six of eight participants regarded Bento box as more efficient, while five of eight participants regarded Bento box as more user-friendly than Oria. Their preference for Bento box mostly depended on their satisfaction with the interface.

Four of five tasks were performed better in Bento box than in Oria. It also required participants to take extra steps to search in Oria than in Bento box even though participants needed to narrow search results at both websites. I observed they were struggling more with Oria than with Bento box, and the total time spent was longer in Oria than in Bento box. Although the five tasks could not represent the whole content search in the system, it certainly showed overall participants performed faster in Bento box than in Oria.

Even known-items search still is a problem for undergraduate students. E-book is the most difficult task to complete. Known-item article search is the easiest task to complete. Open-ended search is difficult for undergraduates. Participants identified unfamiliar features (faceted filters) that they asserted should be retained for their potential utility. To compare findings with other usability tests in library context, this study confirmed some of the previously established findings that undergraduate students preferred a single search box, and only using key words to do search (Kliwer, G. el, 2016). If they used one word to search, the system did not bring

them the pertinent materials in the first search results page, they would come back to the search box and changed to other search word or combination of different words to try again. Undergraduate students usually search for known-items, which develops into a lazy-search habit.

As most usability studies of discovery tools indicate “Google” search behavior affected participants in academic library search (Lown, C., Sierra, T. & Boyer, J., 2013; Hanrath, S.& Kottman, M., 2015; Kliwer, G. el, 2016; Brett, K., Lierman, A. & Turner, C., 2016; Øystby, S.E., 2018; Jaffy, M. 2020). Users like to use a single search box, but never pay more attention to the difference between “Google” and library search with discovery tools, although both have similar appearance with one single search box. “Google” search results do not present many facets menus while in library single search box, finding the pertinent materials mostly depend on using suitable facets to refine the search results.

Undergraduate students have the competence to learn facets quickly if they need to and they are willing to do it. Traditionally, librarians like to simply ascribe the failure of using library’s services to students’ bad information competence. This study showed most participants were struggling with facets whether in Bento box or in Oria. However, most participants could successfully do searching with facets after being prompted. Their behavior changed quickly after the Bento box test and their search strategy got better in Oria. They remembered to use facets filters in Oria. *“An encouraging though unintended outcome of this study was that several participants learned about unfamiliar resources or services”* (Haggerty, K. C. & Scott, R. E. 2019, p.307). This finding warns us: Bento box may not solve all the problems that students felt frustrated with in the Oria even though Bento box may be more user-friendly. Once they needed to do open-ended search, which required them to learn how to refine search results, they would be feel confused as they complained about difficulties of using facets menu in Oria. This finding confirmed what Singley worried about:

*Bento box display doesn’t remove the “silo” problem – library resources still reside in multiple platforms. Users will most likely not find what they are looking for on the “bento” results page, and will need to click “more results” for a particular platform (catalog, article index, library website, etc.). Once in that “silo,” they may not come back to explore other sources.* (Singley, E., 2016)

Therefore, learnability should be a more important factor to evaluate usability of discovery tools to improve academic library services.



## 6.2 Limitations

There are some limitations, although I don't think it would make the findings invalid or make a big difference, I would like to point them out.

### 6.2.1 No screen Recording May Lose Some Information

I did not record audio, screen display, keystrokes and mouse clicks, although the software can supply this function. Due to my experience of being journalist for many years, I could take better notes when I interviewed. The think aloud form helped me record information more efficiently. I have written down the key words and most important information immediately after each test, so I could make up for the lack of recordings and avoid personal privacy issues. However, if I did such a test again, I would ask more people for help. I would plan to use video recording early in the project and apply for permission from NSD. It would be more practical and easier to transcribe test recordings which would be more useful for analyzing the test results.

This test is an online video meeting test, which is different from real "face to face" test. The test participants' interactions with the prototypes were observed via Zoom "shared screen", which means I could not see participants' feelings while they were searching.

### 6.2.2 Need More Participants to be a Proof

Krug confirmed that "*Do-it-yourself usability tests*" aims to improve the current situation by identifying and fixing usability problems. The results based on data collected by qualitative method cannot be simply used as a proof (Krug, S., 2014, p.119).

The data gathered via observation and interviews in this study were most useful in establishing a general picture of undergraduate preference for Bento box or Oria and their reasoning. Although the results showed that most participants had a strong preference for Bento box, it should not be taken simply as proof for changing Oria's current interface. The findings cannot be generalized because eight participants may not necessarily to represent the overall user population, which includes not only undergraduate students, master students, but also researchers, librarians and other users. Participants were from different disciplines, in different universities or colleges, and they were all undergraduate students. Most importantly, I have seen enough search behavior from the eight participants. It does provide information on what undergraduate students think and, more importantly, how they act when searching for various

types of information on the different library's website. If libraries want to take it into consideration, they should do such a test on a larger sample.

### **6.3 Contribution and Further Study Suggestions**

This study is for the first time to attempt to compare Bento box style interface and current Oria interface. The discussion of “To Bento” or “Not to Bento” under the discovery system with Oria – (Norwegian version of Primo) using a prototype test can give an overview of what participants experienced in both websites. The findings suggest that the potentials users' search behavior can be affected by the Bento box. The results have only uncovered the tip of the iceberg. Considering Bento box is quite new in Norway, and it has been used in other countries for years, there are many suggestions for further study.

Tabs (Article, Database, Book, etc.) on the homepage near the single search box were unnecessary, most participants did not pay attention to them. Marc Jaffy (2020) observed their “search journals” tab on their homepage was unnecessary and removed it after redesigning the homepage. This means tabs on the homepage may be unnecessary for most users whether they are placed above or under the single search box.

#### 6.3.1 Recommendations for Further Usability Test

Academic libraries that want to implement a Bento box style design can use my study results to carry out a prototype test or beta test with more participants, or through an online survey to get more feedback. During the coronavirus outbreak period, it is extremely difficult to recruit test participants. I did not want to break the natural user behavior too much because I did not want to increase their stress. If academic libraries decide to provide Bento box search results, they may take into consideration the factors and problems which this study identified through testing.

I learned several lessons relating to prototype testing that would be of value to other researchers seeking to perform similar tests at their own institutions. For example, how far participants' patience would be likely to extend for an online video interview, how many tasks should be applied in the test, what kind of tasks would be better for test, etc. “What remains to be explored is whether students understand the types of resources they are searching, and whether they are using the available features in Primo to best complete their research task” (Kliwer, G. el, 2016,

p.576). Search suggestions like search with “author” “title” “key words” placed in the single search box in Oria, could provide both limits and opportunities for explanation.

#### a. Task Design

Participants in this study did not show they would use library’s website to search things like library website contents, e.g. open hours, borrow or renew books, etc. Bento box promised that users can search both library collections, subscriptions and library website results just through a single search box, allowing users to type things like “how to renew a book”. Consequently, a future test should include such tasks.

Each test provides insights and possible solutions to usability issues, which may help libraries to improve the user experience. Integrating tasks into wireframes makes the testing environment less stressful. If participants did not complete the task sometime, they would be prompted to finish the task before going to the next task, which emphasized searching has different methods. It was not only a test process, but also a learning process.

It was beyond the scope of this study to measure efficiency of the full-text links outside of the library. Further testing, may use a complete version prototype with tasks to find the physical book shelf location in the library and to download e-books or full-text articles. Bento Box style search results display can have shortcomings: for example, causing slow response times. Aaron Tay and Feng Yikang supplied a good technical example for testing Bento box with low cost and less time. They confirmed that “*LibGuides v2 has provided us with a low-cost and effective platform to implement a Bento prototype that allows us to present users with more resources at one time compared to a typical discovery service*” (Tay, A. & Yikang, F., 2015).

Participants did show a strong preference for Bento Box style search results display interface, however, it is worth doing a test based on institution’s own website and recruiting participants from own institution to collect enough data before deciding to replace the older interface. This test basically compared the University of Oslo Library’s interface with the Princeton University Library’s interface. Bento box style display can have different categories. Oria in Norwegian universities and colleges have different interfaces according to their own customizability.

#### b. Using a Think-aloud Form for Observation Note

After completing each online video test, I reviewed the interview notes and observation think aloud form. I immediately wrote down the most important information from observation and interview. I needed to do both observation and take notes at a high tempo, and respond quickly to questions, because I did the testing all by myself. That was a challenge. However, thanks to the think-aloud form I prepared for observation note taking, recording test notes was efficient and useful. For libraries, short of personnel and financial support, this would be a cheaper way to do a usability test. Without an experimental lab and professional teams, it is still possible to identify the most vital usability problems using a discount usability test.

### c. Test Order Matters

Participants' familiarity with Oria website affected their opinion. Two participants who preferred Oria answered that familiarity was a factor in their preference. Further study may consider the test order to avoid possible bias, if they wanted to compare two different systems on the condition that one of which participants may be familiar with before test.

Since familiarity may influence the judgement and acceptance of the new interface, further study may consider whether users' familiarity influences their acceptance of new interface. In my study, one participant strongly requested that we not change Oria interface to Bento box.

### 6.3.2 Card Sorting

Article search was regarded as the most important academic library's service in current discovery systems. Bento box was regarded as the possible best solution for one united single search of the whole library content including articles. As Rochkind concluded, "'bento style' search provides, at present, the best cost/benefit way to improve article search relatively quickly with relatively large user benefit" (Rochkind, J. ,2012). However, this study did not find support that Bento box was especially better for article search than book search. There were some uncertainties. Only two tasks of article search with eight participants can not verify the truth. This study can only shed a little light on the surfaces of possible problems. A big challenge was in gathering data by using Zoom and Balsamiq. It took time to download the tests and connect with the test participants. Sometimes it took even longer than estimated and planned, which made the testing atmosphere more nervous. Though I had hoped for a larger pool of data and a natural atmosphere, I only could get enough responses to draw some tentative conclusion.

Card sorting has not been applied widely to the evaluation of library usability testing. “*Task scenario studies found that users struggle with library jargon and show a preference for clear and uncluttered interfaces*” (Guay, S., Rudin, L. & Reynolds, S., 2019, p 89). For future study, card sorting methods can be recommended for collecting information for designing facet structure. Card sorting and transaction log analysis can reveal insights regarding how to best arrange categories on the interface, how to arrange the faceted features to meet user searching preference and information needs.

### 6.3.3 Students’ Information Competency and Library Information Literacy Lectures

These issues remind us how to measure students’ information competency and how to carry out library information literacy lectures. There are lessons we learned: faceted factors need to be introduced as early as possible; students want to learn the terminologies and how to use the faceted functions; teach the students how to refine their topics and identify keywords for searching. If students want to discover more sources by themselves, for example searching open-ended recourses, not just limited to the known-items, they need to know how to refine the overwhelming search results. There is no shortcut. “*A large number of studies indicated that instruction on Primo or other discovery systems was needed for users to search effectively*” (Brett, K., Lierman, A. & Turner, C., 2016, p.9) Paradoxically, no participants expressed a desire for advanced search options. Their over-reliance on Google search habits have trapped the library’s single search box into a dilemma: have a “Google” like single search box, but it cannot supply all the types of information that users expect. The findings support what Fawley and Krysek (2012) suggested for Library information literacy lectures to “focus on developing search terms” and “teach limiters or facets” (p. 211–212).

We must clarify the benefits of using Bento box search interface before talking about its contribution to discovery systems. Bento box search interfaces appeared in the academic library due to the ambition of integrating the whole library into a single box search while improving article search made it stand out from many discovery tools. It can not only supply the library with a unified single search box but also can compete with commercial discovery tools. In this way, academic libraries be more independent from the commercial vendors with possibly lower cost.

## References:

- Azadbakht, E. S. (2015) Information Literacy Instruction with Primo. *Reference & User Services Quarterly*, vol. 54(3).
- Balsamiq (2020). <https://balsamiq.com/wireframes/cloud/docs/intro/#what-is-balsamiq-cloud> retrieved 15.05.2020
- Beaulieu, M., Robertson, S. & Rasmussen, E. (1996). Evaluating interactive systems in TREC. *Journal of the American Society for Information Science*, 47(1), 1996, 85–94.
- Beswick, K. (2017). *Where We're Headed with Bento-box Search* <http://kevinbeswick.com/bento-box-c4lse2017/#/6> retrieved 05.05.2020
- Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, Vol. 56 Issue: 1, pp.71-90, <https://doi.org/10.1108/EUM0000000007110>
- Borlund, P. (2003) The Concept of Relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10) p.913-925.
- Breeding, M. (2007). *Chapter 5: Primo. Library Technology Reports*. <https://www.journals.ala.org/index.php/ltr/article/view/4541/5335> retrieved 27. 05. 2020 ; *Library Technology Reports*, July-August 2007, Vol.43(4), p.28(5)
- Breeding, M. (2018). Index-Based Discovery Services: Current Market Positions and Trends, *Library Technology Reports*, vol. 54, no. 8. DOI: <https://doi.org/10.5860/ltr.54n8>
- Breeding, M. (2014). Major Discovery Product Profiles. Breeding. *Library Technology Reports*, Vol. 50, Iss. 1, (2014): 33-52.
- Brett, K., Lierman, A. & Turner, C. (2016). Lessons Learned: A Primo Usability Study. *Information Technology & Libraries*, 35(1), 7-25. doi: 10.6017/ital.v35i1.8965
- Bull, S., Craft, E. & Dodds, A. (2014). Evaluation of a Resource Discovery Service: FindIt@Bham. *New Review of Academic Librarianship*, 20(2), 137-166. doi: 10.1080/13614533.2014.897238
- Dabin, E. & Preminger, M. (2018). Discovery Systems as an Alternative to Stand-Alone Databases: The Example of Primo at BI Norwegian Business School. *NORDIC JOURNAL OF INFORMATION LITERACY IN HIGHER EDUCATION*, 2018, 10(1), 17. doi: 10.15845/noril.v10i1.274
- Daul, M. (2014) *Bento is Coming!* <https://blogs.library.duke.edu/bitstreams/2014/07/03/bento-is-coming/> retrieved 04.05. 2020.
- Dempsey, L. (2013). Full Library Discovery. <http://orweblog.oclc.org/archives/002214.html>

- Dempsey, L. (2014). *Network Reshapes the Library: Lorcan Dempsey on Libraries, Services and Networks*. American Library Association.
- Dempsey, M. & Valenti, A. M. (2016). Student Use of Keywords and Limiters in Web- scale Discovery Searching. *Journal of Academic Librarianship*, 42(3), 200-206. doi: 10.1016/j.acalib.2016.03.002
- Fawley, N. E., & Krysak, N. (2012). Information literacy opportunities within the discovery tool environment. *College & Undergraduate Libraries*, 19(2-4), 207-214. <http://dx.doi.org/10.1080/10691316.2012.693439>.
- Fox, V. B. (2017) Cataloging News. *Cataloging & Classification Quarterly*, 55:7-8, 671-684, DOI: 10.1080/01639374.2017.1366159
- Fry, A. & Rich, L. (2011) Usability Testing for e-Resource Discovery: How Students Find and Choose e-Resources Using Library Web Sites. *The Journal of Academic Librarianship*, 37(5), pp. 386-401
- Goker, A., & Davies, J. (Eds.). (2009). *Information retrieval: Searching in the 21st century*. John Wiley & Sons.
- Guay, S., Rudin, L. & Reynolds, S. (2019). Testing, testing: a usability case study at University of Toronto Scarborough Library. *Library Management*. Vol. 40 No. 1/2, pp. 88-97. <https://doi.org/10.1108/LM-10-2017-0107>
- Hanrath, S. & Kottman, M. (2015). Use and Usability of a Discovery Tool in an Academic Library. *Journal of Web Librarianship*, 9:1, 1-21, DOI: 10.1080/19322909.2014.983259
- ISO 9241-11:1998 *Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability*, <https://www.iso.org/standard/16883.html> retrieved 28.05.2020
- Haggerty, K. C. & Scott, R. E. (2019) Do, or Do Not, Make Them Think? : A Usability Study of an Academic Library Search Box. *Journal of Web Librarianship*. 13:4, 296-310, DOI: 10.1080/19322909.2019.1684223
- Hanrath, S. & Kottman, M. (2015) Use and Usability of a Discovery Tool in an Academic Library. *Journal of Web Librarianship*, 9:1, 1-21, DOI: 10.1080/19322909.2014.983259
- Hearst, M. A. (2013). *Search User Interfaces*. Cambridge University Press.
- Hwang, W., & Salvendy, G. (2010). Number of people required for usability evaluation: The 10±2 rule. *Communications of the ACM*, 53, 130-133. <https://dl.acm.org/doi/pdf/10.1145/1735223.1735255>

- Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval*, 3(1–2), 1-224. doi: <http://dx.doi.org/10.1561/15000000012>
- Kliwer, G., Monroe-Gulick, A., Gamble, S. & Radio, E. (2016). Using Primo for undergraduate research: a usability study. *Library Hi Tech*, Vol. 34 No. 4, pp. 566-584. <https://doi.org/10.1108/LHT-05-2016-0052>
- Krug, S. (2014). *Don't make me think, Revisited A Common Sense Approach to Web and Mobile Usability*. New Riders.
- Kuniavsky, M. (2012). *Observing the user experience: a practitioner's guide to user research*. Elsevier.
- Lewis, C.H. (1982). *Using the "Thinking Aloud" Method in Cognitive Interface Design*. IBM TJ Watson Research Center, Yorktown Heights, NY. [https://domino.research.ibm.com/library/cyberdig.nsf/papers/2513E349E05372CC852574EC0051EEA4/\\$File/RC9265.pdf](https://domino.research.ibm.com/library/cyberdig.nsf/papers/2513E349E05372CC852574EC0051EEA4/$File/RC9265.pdf)
- Lewis, J. R. (2014) Usability: Lessons Learned ... and Yet to Be Learned. *International Journal of Human-Computer Interaction*, 30:9, 663-684, DOI: 10.1080/10447318.2014.930311
- Lierman, A., Scott, B., Warren, M., & Turner, C. (2019). Testing for Transition. *Information Technology and Libraries*, 38(4), 76-97. <https://doi.org/10.6017/ital.v38i4.11169>
- Lown, C., Sierra, T. & Boyer, J. (2013). How Users Search the Library from a Single Search Box. *College & Research Libraries*, 74(3), 227-241.
- Mason, J. (2002). *Qualitative Researching* (2. utg.). London: Sage.
- Mischo, W., Norman, M., & Schlembach, M. (2018). Innovations in Discovery Systems: User Studies and the Bento Approach. In B. R. Bernhardt, L. Hinds, & L. Meyer (Eds.), *What's Past is Prologue: Proceedings of the 37th Charleston Conference*, (pp. 299-304). Purdue University Press. <https://doi.org/10.5703/1288284316700>
- Nichols, A., et al. (2014). Kicking the Tires: A Usability Study of the Primo Discovery Tool. *Journal of Web Librarianship*, 8(2),172-195. doi: 10.1080/19322909.2014.903133
- Nielsen, J. (2000). *Why you only need to test with 5 users*. <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> retrieved 12.05.2020
- Nielsen, J. (2004). *Card Sorting: How Many Users to Test* <https://www.nngroup.com/articles/card-sorting-how-many-users-to-test/> retrieved 09.10.2019



- Nielsen, J. (2012). *Usability 101: Introduction to Usability*  
<https://www.nngroup.com/articles/usability-101-introduction-to-usability/> retrieved  
 12.05.2020
- Niu, X., Zhang, T. & Chen, H.-l. (2014). Study of User Search Activities with Two  
 Discovery Tools at an Academic Library. *International Journal of Human–Computer  
 Interaction*,30(5), 422-433. doi: 10.1080/10447318.2013.873281
- O’Connell, R. (2019). Beyond the Bento – A New Discovery Experience for Smith College.  
*Computers in Libraries*. retrieved 06.10.2019  
<http://www.infotoday.com/cilmag/mar19/OConnell--Beyond-the-Bento-A-New-Discovery-Experience-at-Smith-College.shtml>
- Perfetti, C. (2001). *Eight is not enough*. [https://articles.ue.com/eight\\_is\\_not\\_enough/](https://articles.ue.com/eight_is_not_enough/)  
 retrieved 12.05.2020
- Perrin, J. M., Clark, M., De-Leon, E., & Edgar, L. (2014). Usability Testing for Greater  
 Impact: A Primo Case Study. *Information Technology and Libraries*.33(4), 57-66.  
<https://doi.org/10.6017/ital.v33i4.5174>
- Prommann, M.& Zhang, T. (2015). Applying Hierarchical Task Analysis Method to  
 Discovery Layer Evaluation. *Information Technology & Libraries* 34, no. 1 (2015): 77–105,  
 doi:10.6017/ital.v34i1.5600.
- QuickSearch (2020). <https://www.lib.ncsu.edu/projects/quicksearch> retrieved 05.05.2020
- QuickSearch Usability Test (2015). <https://www.lib.ncsu.edu/projects/quicksearch-usability-test>  
 retrieved 05.05.2020
- Rochkind, J. (2012). *Article Search Improvement Strategy*  
<https://bibwild.wordpress.com/2012/10/02/article-search-improvement-strategy/> retrieved  
 04.05. 2020.
- Rose-Wiles, L. M. & Hofmann, M. A. (2013). Still Desperately Seeking Citations:  
 Undergraduate Research in the Age of Web-Scale Discovery. *Journal of Library  
 Administration*. 53(2-3), 147-166. doi: 10.1080/01930826.2013.853493
- Shapiro, S. (2018). Academic Librarians, Information Overload, and the Tao of Discovery.  
*The Journal of Academic Librarianship*, 44 (2018) 671–673.  
<https://doi.org/10.1016/j.acalib.2018.07.004>
- Singley, E. (2016). *To Bento or Not to Bento – Displaying search results*  
<https://emilysingley.net/usablelibraries/to-bento-or-not-to-bento-displaying-search-results/>  
 retrieved 04.05. 2020.
- Spencer, D. (2009). *Card Sorting: Designing Usable Categories*. Rosenfeld Media.

- Snyder, C. (2003). *Paper prototyping: The fast and easy way to design and refine user interfaces*. Morgan Kaufmann.
- Taylor, R. S. (1968). Question-Negotiation and Information Seeking in Libraries. *College & Research Libraries*, 29(3), 178-194. doi: 10.5860/crl\_29\_03\_178rktoy-en- casestudie
- Tay, A. & Yikang, F. (2015) *Implementing a Bento-Style Search in LibGuides v2*. *code4lib Journal. Research Collection Library*. <https://journal.code4lib.org/articles/10709#note1> retrieved 04.05.2020
- Tidal, J. (2017) One Site to Rule Them All, Redux: The Second Round of Usability Testing of a Responsively Designed Web Site. *Journal of Web Librarianship*, 11:1, 16-34, DOI: 10.1080/19322909.2016.1243458
- UNIT (2020). [https://bibsyst-almaprmo.hosted.exlibrisgroup.com/primo-explore/search?vid=BIBSYS&lang=en\\_US](https://bibsyst-almaprmo.hosted.exlibrisgroup.com/primo-explore/search?vid=BIBSYS&lang=en_US) retrieved 03.05.2020
- University of Oslo (UiO) (2020). <https://www.uio.no/english/about/facts/figures/index.html> retrieved 03.05.2020
- van den Haak, M. J., & de Jong, M. D. T. (2003). *Exploring two methods of usability testing: concurrent versus retrospective think-aloud protocols*. 285-287. Paper presented at IEEE International Professional Communication Conference, IPCC 2003, Orlando, United States. <https://doi.org/10.1109/IPCC.2003.1245501>
- Varnum, K.J. (2016) *Exploring discovery: the front door to your library's licensed and digitized content*. American Library Association.
- Wildemuth, B. ed. (2017). *Applications of Social Research Methods to Questions in Information and Library Science*. Libraries Unlimited.
- Wilson, T. D. (2000). Human information behavior. *Informing science*, 3(2), 49-56.
- Yang, S.Q. & Wagner, K. (2010) Evaluating and comparing discovery tools: how close are we towards next generation catalog? *Library Hi Tech*, Vol. 28 No. 4, pp. 690-709. DOI 10.1108/07378831011096312
- Østbye, S. E. (2018). *Studenters bruk av Oria: En kvalitativ undersøkelse av informasjonsbehov i møte med et discovery-system*. OsloMet.

## **Appendix**

### **A1 Invitation letter**

Hei

I am master student at OsloMet. I am going to do a test about library's website search and use to improve the library's search service. I need some undergraduate students to help me do the test. I will send links to participants and observe the search process via Zoom. The test will take about 15-30 minutes.

Thanks for your help!

If anyone could help me, please send e-mail to [xxx@gmail.com](mailto:xxx@gmail.com)

For Norsk

Hei

Jeg er master student ved OsloMet. I forbindelse med min masteroppgave, skal jeg undersøke bibliotekets søkesystem for å forbedre søketjenester. Jeg trenger noen studenter (Bachelor studier) til å delta i et søkeprosess-eksperiment. Jeg skal sende linken til deltakere og observere søkeprosessen. Det tar ca. 15-30 minutter. Testen skal gå via Zoom video Chat.

Tusen takk på forhånd!

Hvis noen har lyst til å hjelpe, vennligst send mail til [xxx@gmail.com](mailto:xxx@gmail.com).

### A2 Think Aloud Form

A:	Step1	Step2	Step3	Step4	Step5	Step6	Step7	Step8	Step9
Bento									
Task 1	Title	T- result							
	Author	A- result							
	Keywords	K- result							
Task 2	Title	T- result							
	Author	A- result	A- catalog	A-catalog keyword author	Catalog- book	Catalog- book- year	Year result		
	Keywords	K- result							
Task 3	Title	T- result							
	Author	A- result	A- result2	Author article2007	Article peer	Article peer 2007	Article journal	Article journal 2007	
	Keywords	k- result	k- article						
Task 4	Keywords	result	article	Article year	Peer journal				
Task 5	Keywords?	Article result	Choose thesis	Thesis result					

B: Oria	Step1	Step2	Step3	Step4	Step5	Step6	7	Step8	Step9
Task 1	Title	T-result	R copy	final					
	Author	A-result	format	Book/ printbook/ year	year				
	Keywords	K-result							
Task 2	Title	T-result	r-books /r-year	Material type/ ebbok/ paperbook	Ebook/eng copy	Ebook copy list	Title r copys	Final r/final open access/	Paperbook year
	Author	A-result	A- paperbook/year	A- paperbook year	Author – year- result	author- book- year			
	Keywords	K-result	1879 result						
Task 3	Title	T-result							
	Author	A-result	A-year	year-peer /author peer year	Author- article	Author- Article -year			
	Keywords	result							
Task 4	Coronavirus	result	Peer/article	Peer relevance Article relevance	Peer date/ Article newest				
	Coronavirus peer	result	peer relevance	Peer – relevance- newest					
	Coronavirus article	result	s-relevance	S date- result					
Task 5	Academic library	result	Choose thesis	Thesis final result					

### **A3 Test Introduction**

For English:

Hello!

Thank you for agreeing to participate. I'll be helping you through the process, and taking notes. Before we get started, I'd like to ask you a few quick questions about yourself.

Now we're going to ask you to complete five tasks. This isn't a test, and nothing you do will be the wrong or right answer. I just want to see how you interact with the site and what we can do to make that experience better.

Do you have any questions so far?

Let's begin. Please try to talk about what you're doing as much as possible, and tell me what you're thinking and why you're taking each step.

Hei

Takk for du vil delta denne testen. Jeg skal være sammen med deg mens du gjør oppgaver. Vi kan kommunisere hver andre underveiens. Oppgavene har ingen fasit. Det handler ikke om riktig og feil. Det er flere søkemuligheter. Ikke redd å gjøre noe "feil". Hvis det er noe uforståelig eller uklar, bare si ifra, så kan jeg fikse det med en gang. Målet er å se hvordan søketjenester fungerer og forbedrer den. Jeg skal observere og ta notater av søkeprosessen via din skjerm.

Nå skal du gjøre fem søkeoppgaver. Dette er ikke prøve. Ingen fasit. Det handler ikke om riktig og feil. Det er flere søkemuligheter. Ikke redd å gjøre noe "feil". Hvis det er noe uforståelig eller uklar, bare si ifra, så kan jeg fikse det med en gang. Jeg håper du kan fortelle hva du skal gjøre og hvordan du tenker på hvert steg.

## A4 Pre –interview Guide

Før testen, vil jeg spørre deg noen generelle informasjoner.

Hvilket år går du nå?

Hvilket fag tar du?

Er du fulltid eller deltid?

Hvor ofte bruker du biblioteket?

Hvor ofte bruker du bibliotekets webside?

Hva bruker du bibliotekets webside til?

Kjønn, Alder, Utdanningsnivå,

Vi bruker "Catalog" for bibliotekets egne samlinger. Forstår du den? Hvis du ikke kunne finne hva du letter etter, hvordan vil du få hjelp?

For English:

- Are you an undergraduate student?
- What do you study?
- Which year are you in now?
- How often do you use this library?
- How often do you use the Libraries' website?
- About how many hours a week would you say you spend online?
- Have you ever used the Oria before? (if not) Have you ever heard of them?

Gender, Age, Education level.

- We call library's own resources as "Catalog." Does that name make sense to you?
- If you couldn't find what you were looking for on this page, what would you do to find help?

## **A5 After-interview Guide**

Now I'd like to ask you a few follow-up questions.

1. Which one do you prefer to use?
2. Which one do you think more efficient?
3. Which one do you think user-friendly?
4. Which task was the easiest one?
5. Which task was the most difficult one?
6. What did you like about using each website?

Website A \_\_\_\_\_

Website B \_\_\_\_\_

7. What did you dislike?

Website A \_\_\_\_\_

Website B \_\_\_\_\_

8. Did it bother you to have to click to the next page to find the relevant materials?
9. Did it bother you to have to narrow the search results?