



UNF Digital Commons

UNF Graduate Theses and Dissertations

Student Scholarship

2019

Active Analytics: Suggesting Navigational Links to Users Based on Temporal Analytics Data

Jacob Koza

Suggested Citation

Koza, Jacob, "Active Analytics: Suggesting Navigational Links to Users Based on Temporal Analytics Data" (2019). *UNF Graduate Theses and Dissertations*. 892.

<https://digitalcommons.unf.edu/etd/892>

This Master's Thesis is brought to you for free and open access by the Student Scholarship at UNF Digital Commons. It has been accepted for inclusion in UNF Graduate Theses and Dissertations by an authorized administrator of UNF Digital Commons. For more information, please contact [Digital Projects](#).

© 2019 All Rights Reserved



Active Analytics: Suggesting Navigational Links to Users Based on Temporal Analytics
Data

by

Jacob M. Koza

A thesis submitted to the
School of Computing
in partial fulfillment of the requirements for the degree of

Master of Science in Computing and Information Sciences

UNIVERSITY OF NORTH FLORIDA
SCHOOL OF COMPUTING

April 2019

Copyright (©) 2018 by Jacob M. Koza

All rights reserved. Reproduction in whole or in part in any form requires the prior written permission of Jacob M. Koza or designated representative.

The thesis “Active Analytics: Suggesting Navigational Links to Users Based on Temporal Analytics Data” submitted by Jacob M. Koza in partial fulfillment of the requirements for the degree of Master of Science in Computing and Software Engineering has been

Approved by the thesis committee:

Date

Dr. Karthikeyan Umapathy
Thesis Advisor and Committee Chairperson

Dr. Xudong Liu

Dr. Sandeep Reddivari

Accepted for the School of Computing:

Dr. Sherif Elfayoumy
Director of the School

Accepted for the College of Computing, Engineering, and Construction:

Dr. William F. Klostermeyer
Interim Dean of the College

Accepted for the University:

Dr. John Kantner
Dean of the Graduate School

ACKNOWLEDGEMENT

I would like to extend a huge thank you to my thesis advisor Dr. Karthikeyan Umapathy who was a fantastic guide and resource through this research process. I want to thank him for his support, guidance, and encouragement in every aspect. I would also like to thank the members of my committee, Dr. Sandeep Reddivari and Dr. Xudong Liu. Both were extremely helpful in guiding the research evaluation process. I wish to extend a thanks to the University of North Florida Information Technology Services, specifically the Enterprise Systems team, for granting access to the Google Analytics data for the university webpage. This research would not have been possible without their support. Special thanks to my previous managers, Hans Priepke and Marian Watters for supporting me in my effort to attain my master's degree.

I would also like to thank all those who took time to edit my thesis to make sure it was up to academic standards. Finally, I wish to extend an enormous thank you to my family in their support of my pursuit of this degree. Thank you to my mother and father, Sara and Michael for their help and constant encouragement, and to my wife Rudy Koza, for enduring with me through the stress of this process. Without all of you, this would not have been possible, and I am forever in your debt.

CONTENTS

ACKNOWLEDGEMENT	iv
CONTENTS.....	v
FIGURES	ix
TABLES	x
ABSTRACT.....	xi
Chapter 1 INTRODUCTION.....	1
1.1 Overview	2
1.2 Contributions.....	4
1.3 Plan.....	5
1.4 Organization	7
Chapter 2 BACKGROUND AND LITERATURE REVIEW.....	9
2.1 Background	9
2.1.1 Theory Background Topics	9
2.1.2 Implementation Background Topics	14
2.2 Related Work.....	17
2.2.1 Web Usage Mining.....	18
2.2.2 Web Log Mining for Temporal Traffic Data.....	20
2.2.3 Adaptive Analytics	21
Chapter 3 RESEARCH METHODOLOGY	23

3.1 Design Science Research Methodology	23
3.2 Design Science Research Guidelines	23
3.2.1 Design as an Artifact	24
3.2.2 Problem Identification and Relevance.....	24
3.2.3 Design Evaluation.....	25
3.2.4 Research Contributions.....	25
3.2.5 Research Rigor	26
3.2.6 Design as a Search Process.....	26
3.2.7 Communication	27
Chapter 4 ACTIVE ANALYTICS FRAMEWORK.....	28
4.1 Website Improvement Process	28
4.2 Active Analytics Framework Architecture.....	30
4.2.1 Data Extraction	32
4.2.2 Data Store Architecture	33
4.2.3 Client-Side Framework.....	36
4.3 Framework Technologies	37
4.3.1 Cloud Hosting Providers	37
4.3.2 Budget.....	40
4.3.3 Technology Decisions	41
Chapter 5 FRAMEWORK.....	43
5.1 Link Selection.....	44

5.2 Homepage Modifications	47
Chapter 6 FRAMEWORK EVALUATION.....	49
6.1 Evaluation Goals.....	49
6.2 Testing Process	50
6.3 Participants	52
6.4 Institutional Review Board (IRB) Approval	53
Chapter 7 EVALUATION RESULTS AND ANALYSIS	54
7.1 Demographics	55
7.2 Independent Samples t-Test Data	58
7.2.1 Task Completion Time	59
7.2.1.1 Time Taken t-Test Results.....	60
7.2.2.2 Time Taken Summary	68
7.2.3 Navigation Steps Taken.....	68
7.2.3.1 Navigation t-Test Results	69
7.2.3.2 Navigation Summary.....	75
7.3 Survey Responses	76
7.3.1 Survey Response Data	76
7.4 Prior Works on Active Analytics.....	78
Chapter 8 FUTURE IMPROVEMENTS.....	80
8.1 Features.....	80

8.2 Timeframe Analysis	81
8.3 Implementation.....	82
Chapter 9 CONCLUSION.....	84
REFERENCES	86
APPENDIX A NAVIGATION TASKS.....	89
APPENDIX B USER EXPERIENCE SURVEY	91
APPENDIX C IRB DOCUMENTS.....	94
IRB Approval Letter.....	94
VITA.....	96

FIGURES

Figure 1: Current process for editing sites	30
Figure 2: Thesis Framework Diagram	31
Figure 3: Data warehouse schema	35
Figure 4: MongoDB Document Example	36
Figure 5: UNF Homepage without Modification.....	44
Figure 6: API Call to Get Top Links	46
Figure 7: CSS modification code for popular links	48
Figure 8: Select UNF Links with Proposed Modifications.....	48
Figure 9: Framework Testing Procedure	51
Figure 10: Responses by Age	56
Figure 11: Responses by Internet Experience.....	56
Figure 12: Responses by Browser	57
Figure 13: Responses by How Often Users Visit unf.edu.....	57
Figure 14: Time Taken to Complete Each Task	59
Figure 15: Steps Taken to Reach Desired Page	69
Figure 16 Qualitative Survey Responses	77

TABLES

Table 1: Amazon Web Services Pricing	40
Table 2: Azure Cloud Services Pricing.....	41
Table 3: Google Cloud Services Pricing.....	41
Table 4: Top Webpages for Weeks 5-7, 2016	45
Table 5: Time Taken Statistics - Non-Normalized.....	61
Table 6: Time Taken Statistical Values	61
Table 7: Time Taken t-Test Results.....	63
Table 8 Navigation Step Data	70
Table 9: Levene's Test and t-Test Results	71

ABSTRACT

Front-end developers are tasked with keeping websites up-to-date while optimizing user experiences and interactions. Tools and systems have been developed to give these individuals granular analytic insight into who, with what, and how users are interacting with their sites. These systems maintain a historical record of user interactions that can be leveraged for design decisions. Developing a framework to aggregate those historical usage records and using it to anticipate user interactions on a webpage could automate the task of optimizing web pages. In this research a system called Active Analytics was created that takes Google Analytics historical usage data and provides a dynamic front-end system for automatically updating web page navigational elements. The previous year's data is extracted from Google Analytics and transformed into a summarization of top navigation steps. Once stored, a responsive front-end system selects from this data a timespan of three weeks from the previous year: current, previous and next. The most frequently reached pages, or their parent pages, will have their navigational UI elements highlighted on a top-level or landing page to attempt to reduce the effort to reach those pages. The Active Analytics framework was evaluated by eliciting volunteers by randomly assigning two versions of a site, one with the framework, one without. It was found that users of the framework-enabled site were able to navigate a site more easily than the original.

Chapter 1

INTRODUCTION

Data analytics and being “data-driven” are becoming ubiquitous concepts in society with the exponential growth of collected online data (LaVelle et al, 2011). Web analytics is a subset of data analytics concepts specifically looking at maximizing metrics for websites, from user retention, exit rates, or usage trends (Phippen, Sheppard, and Furnell, 2004). The origins of web analytics have their roots in crunching server web logs trying to understand usage patterns. Analyzing the data to identify problems and subsequently adjusting websites is a very manual process, often not feasible for companies with large number of webpages. With the evolution and emergence of tools such as Google Analytics that provide live dashboards and historical reports, understanding the data of a webpage has become easier than ever, but updating webpages in response is still a manual process. Research has been conducted in developing automated tools to dynamically update webpages based on this data (Carle, 2015). This thesis looks to expand upon current research to provide an automated tool to actively predict usage patterns of webpages and alleviate the need to constantly monitor and manual adaptation of webpages.

1.1 Overview

Currently, most websites have static user interface elements, i.e., interface design elements do not cater to the contextual needs of a user. Contextual web experiences refer to understanding what users are doing when interacting with a site and altering the experience to improve those interactions (Iskold, 2008). Companies, developers, or coders must make a concerted effort to make a change to the website interfaces to meet the demand of a contextual need. It is far more common for these companies, developers, or coders to place a premium on changing the data or the content, that which represents the business value of the website. However, all too often, while data is being updated or added, the user experience is not being retrofitted, either in terms of general aesthetics or, more importantly to this research, usability. Usability can be described as creating an environment that allows a user to move from point A to point B in the easiest, most natural, or least number of clicks possible (Issa and Isaias, 2015). Most websites that are informational in scope rather than transactional, such as university websites, are updated and redesigned after manual analysis of web analytics report. This requires resources, time and effort, to be allocated to understanding usage patterns and react accordingly. Web usability trends offer much of this information to developers, if they only would choose to examine it. Services such as Google Analytics provide detail metrics on how websites are being traversed and maintain a historical record, which can be analyzed and be accessed through their provided APIs.

Other research has been conducted to take advantage of Google Analytics reporting data to update sites. These systems have been designed to highlight the most popular areas at a given time, reflecting a real-time view of page usage. The first individual to search for a topic under that system would find the system lacking in direction. For sites that maintain a consistent traffic to the most popular pages, this is an adequate approach to improving user experience. However, visitors seeking information that are contextualized to calendar year seasons (spring, summer, fall, and winter) tend to access same pages cyclically. Visitors of a university website visit to glean important information quickly, but their desired information may change throughout the year. For example, prior to the beginning of a semester, students might look for campus maps, advising hours, and bill payments, but nearing the end of a semester, they may begin looking for final exam schedules, advisor office hours and locations, or, again, course offerings for the next semester. A system that considers the change in popularity and the slow ramp-up in ranking of these changing desires needs to be developed and be able to predict what people will want and react accordingly.

As a case study, the University of North Florida (UNF) website was chosen to test a framework that will accomplish the task of analyzing past usage trends, predict user navigation, and adapt accordingly. The UNF homepage alone is comprised of over 70 links between menus, images, and navigational page elements. Elements desired by users may be anywhere from two to seven clicks, depending on the ease of navigation and whether they use the search functionality. By anticipating the resources users desire at a given time of year, and visually indicating which links will better direct them to their

locations quicker, the anticipate result is a reduction in clicks and time required to find a given item.

1.2 Contributions

The goal of this research is to develop a framework by which sites can automatically adapt to the usage trends during a given timeframe. By leveraging data collected by third-party systems, Google Analytics for example, developers can programmatically adjust sites without manual intervention. This thesis looks to answer the question on whether such a system can statistically increase the efficiency of user navigation within a functioning site.

Specifically, points of interest include adjusting the navigational elements on webpages to decrease the overall time and navigation steps required to find a desired resource on a website. Frequently on sites with information transfer focus, such as the homepage of a university, navigation elements are placed without any indication to their overall relevance. Based on usability standards on how people read webpages, important links should be placed top left in response to how users read. With a site containing over 70 links, drawing user's attention to important links in other regions can be beneficial.

There is plenty of existing research on how to determine user patterns from different location, server logs, client-side tracking, and systems like Google Analytics.

Predominantly, once this data has been found, a developer must manually make the

adjustments to the site. Some research has been done to dynamically adjust navigation based on the current most popular page within a site (Carle, 2015). This thesis proposes a system that, once implemented, takes away the manual changes needed to be made by developers and focuses on the time context during which a user is accessing the site. The system eliminates the ramp-up time required for a page to reach the current most popular ranking by actively predicting user needs based on previous data. In order to test the hypothesis of improving user navigation, a functional version of the framework was implemented and evaluated.

1.3 Plan

The University of North Florida uses Google Analytics to capture and maintain usage data at various division levels. The enormous number of navigational links provide a sound testbed to implement a framework that looks to reduce navigational complexity, and the wealth of data available in usage patterns will allow for detailed analysis and application. UNF has been collecting Google Analytics data from 2009 and have a record of over 150 million page views; the UNF homepage alone accounts for nearly one third of traffic, close to 50 million page views. The sheer volume of data available allows for customization on how to apply the adaptive framework.

As a proof of concept and implementation of the framework, a mirror instance of the site was created to be able to track users through a series of designed tasks to understand the usage of the site with the new system compared to the original site. Performance was

evaluated based on quantitative metrics such as number of clicks and time spent per task, as well as qualitative research in the form of response surveys where users indicated the level of perceived effectiveness.

The first portion of the framework involves querying the appropriate information from Google Analytics. One benefit of, but also a deterrent to, Google Analytics is the volume of information available; there are over 260 dimensions and over 160 metrics that can be gathered by Analytics, not including deprecated features (Google Analytics, 2017). The amount of data available allows for detailed analysis, but also can be daunting when unsure of what one is truly looking to understand. For the purposes of this research, the number of data points collected from Google Analytics was reduced in scope, limited to dimensions and metrics concerned with views and usage trends, specifically focused on data from the last year.

Once the system to gather the data from Google Analytics was developed, the data needed to be stored in a usable format. The system parses the data from the local data store to extract usage trends for the current timeframe. The working time range was set to a sliding two-week window, a week prior to and a week following the current day. This system analyzes the usage trends for the current timeframe a year prior and nightly summarizes the data in an easily accessed report.

Finally, a client-side interface was created to enable query data from the reporting store. This system was designed to read from the store and dynamically adapt the visual

appearance of the webpage based on the usage trends from the past year. The system was designed to be generic enough to be applied to a variety of web sites, not limited to the UNF site.

1.4 Organization

This paper is structured in nine different chapters. In the second chapter, current research is evaluated in related fields including web analysis and web usage mining, as well as conduct a literature review of different current publications on the subject. This research falls under the purview of the Design Science Research Methodology (Hevner, March, Park, & Ram, 2004), and in Chapter Three this methodology is explained how it was applied it to the current research objective. In Chapter Four the current framework is explained in detail, as well as a description of proposed modifications, including its architecture and how each of the pieces will be implemented. In the fifth chapter, a mirror of the UNF website with two alternate views is created, and the fully developed framework is assigned to one. The strengths and weaknesses of the proposed testing environment is evaluated as well as laying out the foundation for the tests. In the sixth chapter the framework is subjected to various user tests, both on the vanilla UNF website and the site with the proposed modifications, and the interactions are recorded. In the seventh chapter the results of this evaluation are analyzed to calculate statistical measures to determine if the framework has a statistically significant impact on usability. In the eighth chapter, the limitations of the work are discussed, and areas of proposed future research are identified. Finally, Chapter Nine concludes the research with a reflection

response to the initial research hypotheses and the frameworks proposed benefits. There are appendixes attached after Chapter Nine that include various forms that were used during the thesis project.

Chapter 2

BACKGROUND AND LITERATURE REVIEW

2.1 Background

In this section, relevant background information that are the base the proposed framework are discussed. There are three major topics to the research: web analytics and metrics, web usability, and the key topic to this research, the concept of usage timeframes. These topics are discussed at length to understand the current state of the industry. In addition, an in-depth analysis of other relevant topics is provided. Having a deep understanding of the major and tangential concepts above will be critical to the development of the framework.

2.1.1 Theory Background Topics

2.1.1.1 Web Metrics and Analytics

Web metrics record and analyze visitor traffic to and through a web site (Khoo et. all, 2008). These metrics are observable, recordable measurements corresponding to the usage of websites, and can include attributes such as click count, page views, entry point, and bounce rate (Dhyani, Nq, and Bhowmick, 2002). Companies and developers can use these insights to understand the usage patterns of their websites and adjust accordingly. Traditionally, web metrics were used in marketing strategy by companies to maximize

the use, and hopefully subsequently the profit made by their sites, but increasingly companies are investigating the benefits of web metric analysis.

The rise of web usage mining to be able to pull web logs and analyze them for patterns occurred because of the push by marketing companies to maximize profit through web sites. Web usage mining is the process of extracting server logs and analyzing the content therein to identify and understand trends in web usage. Server logs generate information and records based on every interaction between the user and the server, and many of these metrics are stored there (Büchner and Mulvenna, 1998). However, there can be a fair bit of noise in the server logs; information can be included that is not solely the usage behavior of a user, and server logs are restricted only to the information that is passed through the traditional HTTP request. A second form of web usage mining involves the use of client-side scripts or add-ons that monitor the user interactions with a webpage. The benefit of using client-side information gathering is that there may be additional information available such as screen size, or mouse movements that would not be included in a server log (Zheng and Peltsverger 2015).

Web Analytics is the “measurement, collection, analysis and reporting of web data for the purposes of understanding and optimizing web usage” (Burby, Brown, 2007). There are numerous web analytic tools that can be used as either standalone products or are bundled within applications for reporting. New technologies gather and report a wide variety of metrics at different granularities, from individual page views to usage trends, which can be consumed and analyzed by companies to better understand or design their websites.

Other metrics can include information by users and sessions: geographic location, referring website (the page they were on immediately before), and bounce rate (how many people ended their session on that page). Currently the market is heavily skewed towards Google Analytics. As of November 2016, W3Tech market analysis research estimated that 54.9% of all website using Google Analytics, which accounts for 83.5% of the current market share (W3Techs. 2016). The next three smaller tools include Yandex.Metrika (5.3% use, 8.1% market), WordPress Stats (4.5%, 6.9%), LiveInternet (3.3%, 5.0%) and no other tools has more than a 3% market share. Due to a commanding control of market share, free service, and quality of reporting and data features, Google Analytics results will be the primary source of data for this research.

2.1.1.2 Web Usability

Web usability is concerned with the effectiveness, efficiency, and satisfaction that users experience when they are trying to complete any specific task on a website (Geng and Tian, 2015). Additionally, usability incorporates the quality of the service that a user receives from a site as websites are only effective when they are being used. If a user cannot accomplish their task in a normal interaction with a site, it is of no value (Dias et al, 2012). Confusing navigation, haphazard site structure, and even poor aesthetic choices can all be a detriment to web usability. Developers need to consider the ease of use for their site to maximize efficiency and value; in transactional websites, a loss of a user results in loss of revenue, but this can be applied to non-transactional sites as well.

According to research, there are three aspects, or design principles that apply to usability on the web: structural firmness, functional convenience, and presentational delight (Geng and Tian, 2015). These represent the site in a holistic manner. Structural firmness references the strength of the site on matters such as security and performance. Metrics, such as failure rates, 404 Page Not Found errors, and secure connections, influence the usability of a site. Functional convenience deals with the user's interaction with a site: how easy the site is to navigate, intuitive controls, and presence of help documentation. Finally, the aesthetics of the page or site play a crucial role in determining usability, the concept of presentational delight. Using Cascading Style Sheets (CSS) and interactive controls from frameworks such as JavaScript impact the use of a site. From the bottom up, sites need to address the concept of usability to provide a site that attracts users rather than presents barriers to use.

2.1.1.3 Web Personalization

Web personalization is the idea that a website should provide what users want or are looking for without explicitly asking it (Deddegowada et al, 2016). Every person has an online "personality," what topics they are interested in, what sites they frequent most often, and their specific search patterns. Data from disparate sources can be aggregated to present a holistic view of a user; the focus of data mining and data analytics is the combination of heterogeneous data sources to gain some intelligent insight, one example being a user's shopping preferences.

Web personalization can come in two distinct forms: technical or automatic and non-

technical or manual. Web servers and client-side page interaction logging techniques capture most, if not all, of the interaction between a user and a site. In these instances, a user has a digital identity that can be uniquely tied to them. Data mining techniques such as text mining, search mining, the use of cookies, or image mining, and server log mining are the several ways to analyze a web personality, but it requires significant effort to capture and aggregate this data since it comes from distributed sources (McGrath, 2008). A server log may capture a user request history but a client-side method captures more detailed information such as mouse movement.

The second method for personalization requires input from the user to manually enter information (Bhattacharya et al, 2016). The use of ratings can be a critical way to recommend or further personalize the web. Users can evaluate a site's look and feel, ease of navigation, or overall complexity based on a standard point scale. Users can pick what topics interest them; many sites such as Pinterest or StumbleUpon have sections in signing up where users specify their interests. This allows an algorithm to determine possible items to display that match the given criteria. Other sites, such as music streaming sites like Spotify or Pandora allow constant recommendation adaptation by allowing users to manually select whether a song was enjoyable or not. The benefit of the manual approach allows for less errors when trying to match a user's interest as they have specified what they want.

2.1.1.4 Web Timeframes

Website traffic fluctuates over time; commercial sites tend to be more active near major holidays, blogs may see increased traffic around the release of a new article, other remain at a more consistent level year-round (Khoo et al, 2008). Depending on the sample size being analyzed, web metrics can drastically change. A granular investigation in traffic during June may not adequately identify use during peak months for an e-commerce site such as Amazon. Peak months can include November and December for sites such as Amazon, while for flower and chocolate sites may see an increase in traffic around Valentine's day in the United States. At the same time, taking too broad of a sample can remove any nuanced changes that occur in traffic. Looking at an informational site, the top viewed page overall may be skewed as a result of a burst of traffic during a given period and does not accurately reflect the seasonal desires of users. The ability to understand the temporal data of usage trends on sites requires longitudinal data over a year's worth of data would be required to understand seasonal fluctuations in usage (Khoo et al, 2008).

2.1.2 Implementation Background Topics

2.1.2.1 Data Warehousing

A data warehouse is a combination of heterogeneous data sources for the purpose of analytic reporting and direct queries to assist in making business decisions. By nature, data warehouses are subject-oriented, integrated, time-variant, and non-volatile. Data

warehouses are meant to handle large volumes of data, and as such they are the most appropriate way to store the data used for this research. A basic description of data warehouses as they apply to this research is provided here.

Data warehouses can be used to analyze a given subject matter, for example, web page usage for the UNF website will be investigated. Like stated before, data warehouses are integrations of a wide variety of data, often in different, possibly conflicting formats. Data warehouses are focused on historical data and reporting. One of the key fields of records being entered in a data warehouse are date and/or time stamps. This allows for continual historical representation of the data. For this research, historical records of page views for the last eight years with granularity to the day is collected. Finally, data warehouses are non-volatile, meaning that data is only ever entered, never removed. This research will maintain information gathered from Google Analytics and hold it indefinitely so that further research could be done on the historical records.

The first key step to implementing a data warehouse is determining the schema. Two styles of schema have been identified as standards for organizing data: star and snowflake. The star schema is the simplest form of a structure comprised of only two objects: fact and dimension tables. The fact table is the central structure, the representation of the data that is being analyzed in a data warehouse. These records are snapshots in time, due to the time-variant nature of warehouses, and provide top-level details. Many of the properties of a fact table are keys to the dimension tables, which hold more detailed information about each fact records. A snowflake schema is like a star

schema, but it applies the idea of normalization to the dimension tables, spreading the size of the schema. For the system, a simple star schema to hold data information is implemented. More data about this is provided in the section 4.2.2

The second part of data warehousing is the Extract, Transform, Load (ETL) process of data gathering. The goal is to gather heterogeneous data from varying sources: unstructured data such as text files and emails, semi-structured data such as XML or other markup languages, legacy mainframe systems, and edge computing devices (Mukherjee and Kar 2017). Practices such as aggregating, cleaning, and filtering are used to transform data into a more uniform structure to allow for analysis. Finally, data is then inserted, or loaded, into the data warehouse following the schema laid out. ETL processes can be automatically run on schedules or can be loaded in one-off instances. For this research, a custom ETL process was designed to run nightly to gather data, albeit from a single source, more information in section 4.2.1.

2.1.2.2 Web Services

Web services are built upon platforms concerned with the integration of enterprise applications and their data. Each service is an application that delivers some functionality or information over the Internet (Sattiaseelan et al, 2009). They are comprised of open standard interfaces and protocols which allows for identical communication between the service and any application that interacts with it. Many of these services are the fundamental building blocks of larger service-oriented applications ranging from

business information management, to business-to-business sales, and business application integration (Zhang and He, 2008).

Most of today's applications either communicate with, or present their own, Application Programming Interface (API). Sites such as Facebook, Twitter, and Google, have dedicated platforms to communicate with their various servers. Google has a suite of APIs covering YouTube, Gmail, Advertisements, and most importantly to this research, their web Analytics data. These examples are fully-formed, large corporation web services but are more than a sum of their parts. Each of these services is composed of many smaller services that were built over time and integrated into a full platform. These combinations provide a more sophisticated web service for end users to consume (Sathiaselvan et al, 2009). By accessing existing web services, and incorporating them with new or novel technologies, programmers can expand the knowledge base and provide some functionality either to themselves or to others with access. This research is focused on combining existing services with this novel approach to create a flexible framework adaptive to users need. More information about the web service is provided in section 4.2.3.

2.2 Related Work

In researching analytics and personalization of the web the topics of web usage mining and web usability have been well researched and documented. The initial trend in usage mining included server-side log collection and analysis and evolved to include the use of

client-side tracking systems such as Google Analytics. Research into adaptation of web pages based on the usage trends to maximize efficiency is also prevalent. However, there is a lack of research focused on the automatic adaptation of webpages either based on overall trends, or specifically as it relates to this research, how to adapt pages based on identified timeframes. The focus became, then, pertinent literature on the concepts of usage mining and usability since this research was found to be the most influential to the framework development.

2.2.1 Web Usage Mining

Web usage mining has its origins in the collection, synthesis, and analysis of web server logs that websites reside on. Before the introduction of client-side analytical and reporting tools such as Google Analytics data was collected from server-side logs. These web logs were useful because they maintained a record of every interaction between a client machine and the host that occurred over HTTP traffic. The issue then became how to collect, combine, and analyze these data contained within these logs since transactions were stateless and could not easily be tied to individual users (Reddy et al, 2012). The research conducted by Alex Buchner and Maurice Mulvenna proposes a strategy to combine web server logs into a three-dimensional hypercube using a traditional star data warehouse schema (Buchner and Mulvenna, 1998). This proposed system allowed for traditional Online Analytical Processing (OLAP) and data mining techniques. The overall goal of their research was to garner market intelligence, specifically looking at customer attraction, retention, and overall sales. The benefit of using web server logs is the

standardization of logs, either in the standard format, known as NCSA Common log format, or in the newly established Extended format written by W3C in WD-logfile-960323 (W3C, IBM). This standardization allows for research and methodologies to be utilized by a wider audience with little need of customization. Their research does not directly relate to the adaptation of webpages based on usage trends as their focus was on extracting web log data instead of analytics data, but their concept of mining server logs to create an overview of web usage provides a strong foundation for developing a schema and warehouse design when implementing this framework.

Web personalization has become essential to the web to adapt content to visitors' expectations (Nourget et al, 2006). Research by Sajeev and Ramya looks to create a framework to help businesses or developers anticipate what users need and suggest web pages based on base navigational patterns (2016). Their research is a combination of web usage and web content mining, looking to generate a user profile based on browsing characteristics and to suggest relevant pages based on the content of pages they already visit. Using term frequency identification on web pages themselves, combined with extraction of log data to generate sequences of pages, user patterns can be identified and suggested upon (Sajeev and Ramya, 2016). The portion of their research that is crucial to ours is the log data processing to extract usage streams, following a user from page to page to understand their habits. This research is not concerned with web page content but is strictly focused on the result of a user's interaction with a site.

2.2.2 Web Log Mining for Temporal Traffic Data

Web log mining is useful for analyzing the high-level traffic on websites for commercial applications, such as placing ads on the most relevant sites, but these logs lack the fine-grained detail to make meaningful decision on how a site is functioning (Khoo et al, 2008). Web metric tools have gained popularity in producing actionable intelligence on the usage of websites including metrics such as number of visits, page views, or link clicks. These systems rely on client-side browsers loading pieces of code, usually JavaScript, to gather information about the client machine. This leads to an immense amount of data gathered per session. In their research, Khoo et al. looked at applying web metrics to understand digital library assets. In their research, they found that it is difficult to gather and apply metric data since there is a lack of standardization among services (Khoo et al, 2008). Despite the difficulties identified by Khoo and others, web metric data was chosen to base further analysis on as it provides more information on which to segment, specifically the referring pages, subsequent pages, and timeframes. They are one of the few studies that explicitly identify the temporal aspect of web usage, looking at fluctuation in use over a given timeframe, which is one of the cornerstones of this research.

In their research, Khoo et al. identified that web site traffic fluctuates over time (2008). They identified in their research on digital libraries that traffic was lower on the weekend and during holidays, due to the vacation from school by most users. Understanding the nature of the system and potential causes for these fluctuations aids in the understanding

of traffic patterns; a decline in usage in December for a school website, such as UNF's web site, may be an expected occurrence, but a decline in traffic for a company such as Amazon during what should be their peak season may instigate further investigation. Khoo et al. investigated the Library of Congress, National Science Digital Library, and the WGHB Teachers Domain to understand importance of web metrics. In this research they were focused on session times during different months of the year. They found that there were some fluctuations to session length in average use in the different months sampled. The goal is to further this research and focus on usage patterns in web sites to predict future usage in the same time period.

2.2.3 Adaptive Analytics

As discussed earlier, acting on analytics data to update web page interfaces to surface the most relevant information requires manual intervention by developers after reviewing the information. Recent research has been conducted trying to provide an automated solution that would dynamically adapt a webpage based on the most popular pages (Carle, 2015). In that research, upon which this research is based, Carle took web analytics data for a university website and created a framework to process the data and dynamically update the webpages based on the current most popular pages. This was with the intent to surface the pages under the assumption that people looking at the site at the same time would likely be looking for similar things. In their research, where they tested seven different tasks, they found that most of the tasks experienced a reduction in navigational steps and overall complexity to find a given item.

While the foundation of this research is focused on the framework developed by Carle, there are a few key distinctions between the two that need to be highlighted (Carle, 2015). Both sets of research focus on the same site, and the same data, as both had access to the analytics information and the site code. In his research, however, Carle focused on highlighting frequently used navigational menu options in a page based on trends in a week timeframe. This required additional computation and caching to reduce the number of programmatic calls to Google Analytics API. Also, it did not take into consideration the historical usage patterns of any pages. In this research, the computational power needed is reduced as the API call is limited to a nightly process. The focus is also on the historical nature to predict usage, not the current usage. The extension of this research is based on one of the improvements identified by Carle himself, in which he stated “looking back to previous year’s analytics to get a sense of what might become popular in the near future and serve that content up more prominently just as it is needed” could be an improvement (Carle, 2015).

Chapter 3

RESEARCH METHODOLOGY

3.1 Design Science Research Methodology

For this research, the Design Science Research Methodology was used; the goal of such research being to create new or novel technologies, through the creation of new knowledge. In addition to creating new software, it is imperative to analyze the use or performance of this newly created artifact and reflect upon its merits (Vaishnavi and Kuechler, 2015). What this is really illustrating is that the research and creation of a novel artifact is, in fact, a valid and acceptable research methodology. In accordance with this assumption, the goal of this research is to put forth a fully functional product and test it under real world scenarios to understand and quantify its potential impact, whether positive or negative.

3.2 Design Science Research Guidelines

The Design Science Methodology is inherently a problem-solving process; the knowledge and understanding of a design problem comes from the physical building and application of an artifact (Hevner et al, 2004). Guidelines were established to help researchers determine the necessary goals or activities of effective design research. While each of these guidelines should be addressed in some form during the research and design to ensure quality results, they are not meant as mandatory activities to be followed in

order. Researchers have the flexibility to apply these guidelines when and where they deem fit.

3.2.1 Design as an Artifact

As the goal of design science research is the creation of an artifact to address a problem, the goal of this activity is to make sure that the output of the research is a system where the research object is embedded in the design. This includes determining the desired functionality and architecture and taking that to create a physical, working object. In this research both a model and an instantiation of the Adaptive Analytics framework was created, which serves as the artifact against which the effectiveness of the proposed solution is tested. A detailed overview of this artifact is provided in Chapter 4.

3.2.2 Problem Identification and Relevance

The developed system, as an outcome of design science research, should be an innovative solution to an important and relevant business problem so this guideline helps ensure the artifact being designed provides a real solution. For this research the problem identified is the static nature of a website landing page in response to changing user demands over given timeframes. Web traffic is temporal in nature and needs change at different points in time. There needs to be an automated way to provide indicators on a top-level page as to where the most likely resources users are looking for can be found during the identified timeframe.

3.2.3 Design Evaluation

The usefulness, quality, and ability to solve the identified problem of the solution needs to be demonstrated through rigorous and well-known evaluation methods and as such this guideline is focused on determining metrics and analyzing results to show the changes provided by this research. For this research a mirrored version of the UNF website was created in two instances, one with the framework applied and one without for the purposes of evaluating the Adaptive Analytics using blind A/B testing. Users were asked to complete a set of instructions after being randomly assigned one version of the site. Quantitative metrics on the completion of these tasks and qualitative metrics to the users' feelings of overall effectiveness were compared to determine whether the system provided a solution to the problem statement. The research methods are explained in Chapter 5 and in Chapter 6 the data received is evaluated.

3.2.4 Research Contributions

The research contributions guideline establishes that the research must provide clear contributions in terms of the artifact and the knowledge associated, both in construction (foundations) and evaluation (methodologies). This framework provides a generic, novel platform on which users can query historical web usage data, aggregate that data, and adapt their webpages based on given timeframes. It is generic enough that it could be applied to websites that uses web metric data, either from server-side logs, or from third-

party client-side alternatives. The resultant site with framework reacts to historical data to provide adaptive adjustments to appearance to reduce the complexity of the path to a desired resource.

3.2.5 Research Rigor

Design science requires the application of rigorous methods in both construction and evaluation of an artifact and this guideline ensures those methods are met. Rigor is derived from the effective use of the existing knowledge base. Each decision made in the creation of this framework was based on exhaustive review of current and past research to determine the best practices. The assessment approach uses well-documented, universally accepted statistical measures to determine if the framework supports the hypothesis or not. The research rigor is derived from the effective use of current knowledge in the domain space.

3.2.6 Design as a Search Process

This guideline is focused on the idea that design science is an iterative approach to discovering the most effective solution to a problem by using available resources to reach the desired outcome. The framework is a culmination of exhaustive analysis of current and past research in the topic areas of web adaptability, usability, and customization. The research did not begin on a blank slate; existing frameworks were leveraged with other

ideas to solidify the methodology used. The artifact was developed iteratively, at each stage it examined whether the chosen solution would provide the best result.

3.2.7 Communication

Any developed solution does not exist in a vacuum; this guideline focuses on the idea that once an artifact is created it needs to be effectively communicated to both technologically and management-oriented individuals. The artifact needs to be expounded upon in terms of both technological revolution as well as the knowledge required to implement the solution in business. In accordance with this guideline, this research will be presented at a public forum, for both technical and management-oriented individuals, and the source code of the artifact created will be made publicly available.

Chapter 4

ACTIVE ANALYTICS FRAMEWORK

In this chapter, the proposed active analytics framework is discussed and how it was developed based on the design choices made throughout this research. The process began by examining current processes that developers use to update their sites based on analytics data. Following that the architecture of the proposed improved system is presented. First, the Extract, Transform, and Load process used by this framework to move the data from Google Analytics to the selected data store is explained. Following that, the structured of the data warehouse schema used to store the analytic data in a format for consumption by the front-end service is laid out. After, the client-side JavaScript framework used to read from the data warehouse and make the necessary adjustments to the website is developed. This chapter finishes with a discussion on the potential technology platforms available to host the framework and deciding upon the most appropriate technology stack to fit the testing needs.

4.1 Website Improvement Process

Currently, web analytics data typically are store in a separate server from that of a web server where a site may be hosted. In order to use analytics data, developers need to access the data, review it, and make decision about how to change their site to influence the metrics. Figure 1 below represents the traditional sense of web analytics as they relate

to web usage. When accessing a website through a traditional CMS system, content is served to users from the CMS server. Traditional architecture can be set up so the application sits on a separate server from the database server, or both can be installed on one server. Script embedded on these web pages gather information about a user's session and send that data to a data store. In the instance of Google Analytics, data is transmitted from the browser to Google's servers. In order to update pages based on any analytics data, the developers must access the data, wherever it is located, and analyze the results. Based on these results, the developers must log back into the CMS system and make manual changes to the sites to address the trends shown in the data. Subsequently, developers must wait until more people access the web pages to gather additional data for future changes in an endlessly cyclical interaction. This process is non-automatic; it requires data to exist before changes can be made, and all changes must be done manually. Even with the abundance of tools and automation, this process has become taxing and obsolete.

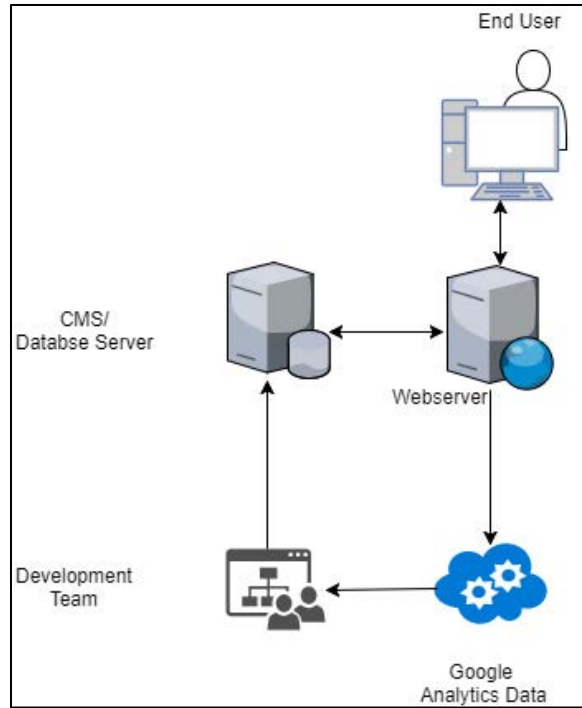


Figure 1: Current process for editing sites

4.2 Active Analytics Framework Architecture

The process outlined above has been the de facto way to adjust websites based on analytic and usage data in the past, requiring significant time investment for each change. The framework provides an alternative, automated way to adjust websites based on changing data patterns. This removes the need for hands-on changes to the smaller minutiae of a site and allows for higher-level discussions and changes. Instead of a developer or team gathering analytics data and applying business knowledge to the metrics to determine change, this framework serves as the agent of change. It reads information directly from the analytics data store, processes the data to match a defined schema, and then automatically applies these changes as interface alterations to a site.

While it does not remove the need to data to exist, it removes the need for manual intervention. Developer effort is needed to understand the system and organize the schema, but then the framework becomes an autonomous agent for altering a site. This system requires a solution that queries data directly from Google Analytics and converts that information into a form that is then hosted by the second major portion of this framework, the data store. Finally, there is a client-side component that is the consumer of the data in the data store; it takes the information based on the business logic of the data store and applies user interface changes to the site. Figure 2 depicts the structure of the proposed framework.

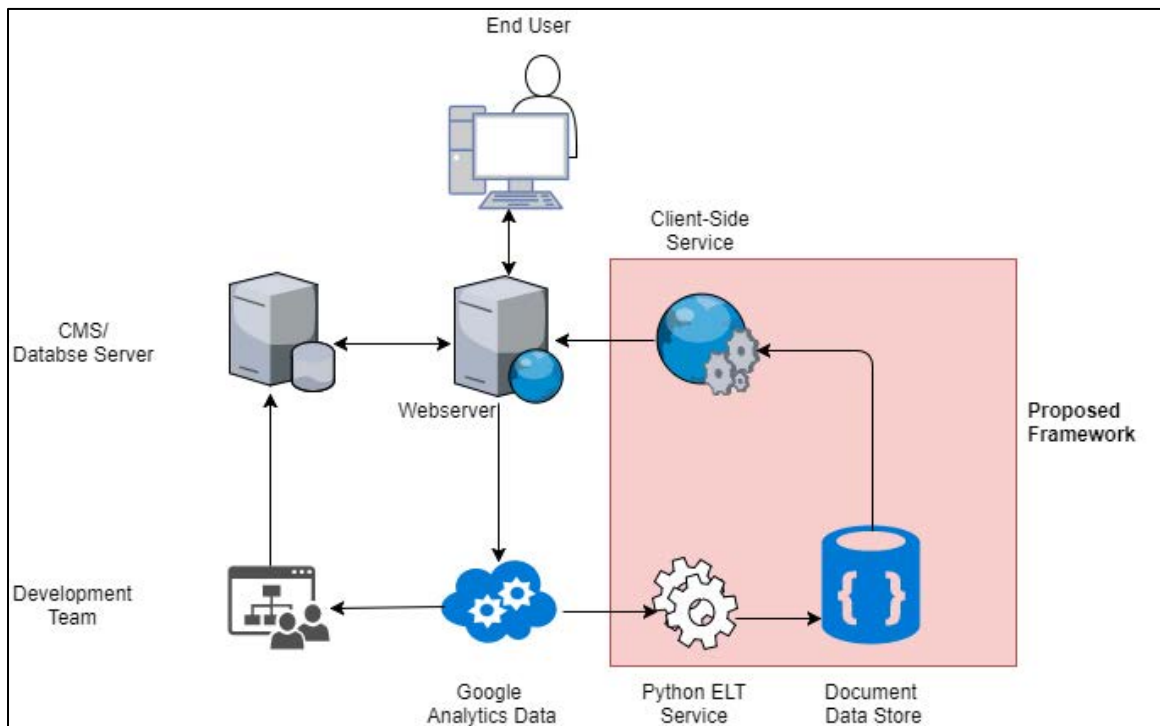


Figure 2: Thesis Framework Diagram

4.2.1 Data Extraction

The first piece of the framework involves getting the usage data from the analytics data provider. Google Analytics was chosen for its robust feature set and since it is the primary analytics data store for the site it was chosen as the testbed for the framework. Google Analytics holds an extreme portion of the market share; nearly 70% of all sites that use an analytics service use Google Analytics, which is nearly 65% higher than the next leading competitor (Datanyze, 2017). Google Analytics provides tracking codes that should be embedded within a site that allows it to gather user actions on the site as analytics data. Developers can access gathered analytics data from the Google Analytics site. Google Analytics also provides a suite of RESTful APIs for accessing data gathered programmatically. Each of these APIs are rate limited services which would limit the amount of endpoint calls available.

Two drawbacks exist when trying to use Google Analytics for data gathering that must be addressed: rate limiting and its focus on reports. Google Analytics is a free service, but to prevent misuse or overuse, each of the API endpoints is rate-limited by account, restricting the number of requests per minute. Since the framework is not reliant on real time data but, rather, focused on past usage information, there is flexibility to space out calls to the API to return data. Also, Google Analytics is focused on presenting reports of usage to clients, so the data may not be in the format needed to be consumed by a front-end service.

Since the framework only needs to return a yearly capture, it is not limited by the API restrictions presented by Google Analytics. The data is static, historical in nature and for the analysis and modifications, and the ETL process does not need to be run frequently. Once a year, on the first of the year, would provide all the data that was gathered for the previous year. This is beneficial because there is no need to account for rate-limiting in the ETL process. The static nature allows us to pull all the records in one call, and parse and store the data as necessary.

The data pulled from Google Analytics needs to be formatted to match the desired data store schema. Data provided by the REST APIs is returned from a single endpoint; sets of dimensions and metrics are passed to the services and a single JSON object is returned. Dimensions represent the attributes of the data, such as pages, while metrics are quantitative measurements such as page views, sessions, or connections. More information on the data store schema is provided in the next section, 4.2.2.

4.2.2 Data Store Architecture

The next portion of the framework deals with storing the information retrieved from the Google Analytics API. The data store design has many similarities to traditional data warehouse design; data warehouses are subject oriented, integrated, time-variant, and non-volatile and are meant to handle large volumes of data and the implemented system incorporates these concepts. The warehouse design was based on statistics regarding the historical popularity of a site during a given period. The information is stored as a single

record in the database corresponding to each URL's specifics in a time variant method. Since the research focuses on the importance of past popularity during structured timeframes, data pulled from Google Analytics focuses on the provided metrics of WeekOfYear and Year as a de facto time stamp. Using both metrics allow for variability in selecting subsections of the data for varying granularity.

The data store design follows the structure of a traditional star schema used in data warehousing with fact and dimensions tables (see Figure 3 below). The schema is designed based on the foundational time-variant nature of data warehouses. The focus of the research is to determine if using analytic data from previous year(s) for a given timeframe could provide better overall site navigation. As such the core of the fact table is the use of the WeekOfYear metric from Google Analytics. This metrics is a numeric representation of the week of the year the data was collected on. To make the framework extensible, the year field was included as well, so that if data is available for multiple years, developers could alter the framework to combine these data points for a more comprehensive analysis. In the dimension tables additional details are provided to allow for further examination on both the timeframes and the pages, including the specific date range a week corresponds to or supplemental page information such as average time on page or exit rates. These allow for further customization of the overall framework giving the power to the developer to decide what supplemental information is useful for their site. Popularity can be viewed in terms of generic page views or exit rates can be used with the hypothesis that additional exits per page suggest users found what they were looking for.

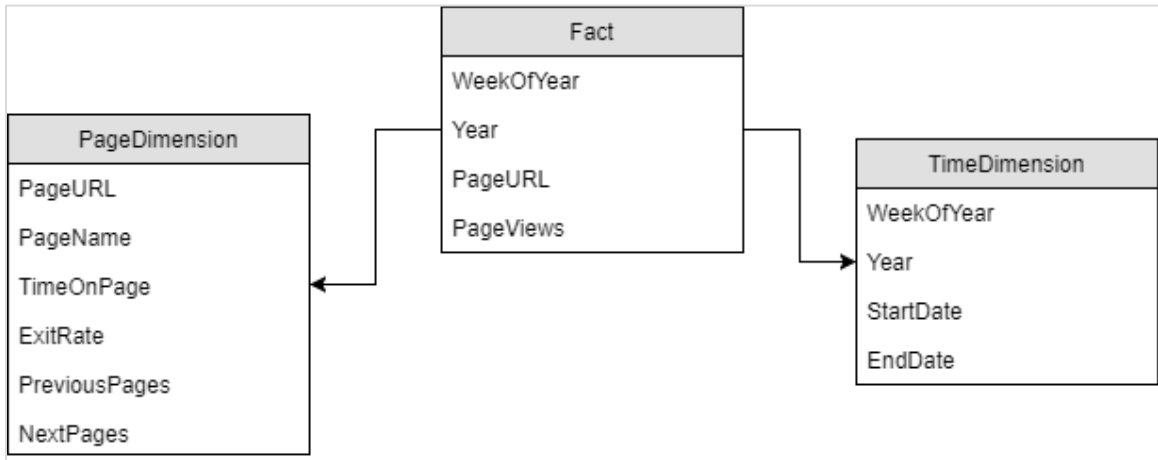


Figure 3: Data warehouse schema

Due to the sheer volume of data that will need to be queried and added to the data store for information retrieval, NoSQL databases will be leveraged to store the data. NoSQL can come in a variety of formats, including document databases, graph stores, key-value stores, and wide-column stores. Since the data is static, it is historical data that is being queried, the document store schema is chosen for its simplicity and ability to return JSON elements. A document database would hold a collection of documents which represent the real dataset, like a collection of rows in a traditional relational database. Specifically, MongoDB is used as the NoSQL data store since it is a free, open-source alternative that can be run on any platform. It allows for quick retrieval of documents stored in JSON format, allows for ad-hoc queries, and is focused on being highly-available easily distributable at its core (MongoDB). In this instance, collections are broken out by week of the year. Under each week collection, every document represents a single page dimension. Figure 4 below shows an example document with the data stored.

Key	Value	Type
(1) ObjectId("59d5701d5eab5718bc910f08")	{ 4 fields }	Object
_id	ObjectId("59d5701d5eab5718bc910f08")	ObjectId
url	/default.aspx	String
isoweek	04	String
pageviews	96307	Int32

Figure 4: MongoDB Document Example

4.2.3 Client-Side Framework

The last piece of the framework is the client-side portion that updates the look and feel of a website. This component adapts the user interface of the top-level sites by altering the appearance of links that are associated with the URL in the data store. Links that are more popular during the given timeframe need to be visually different from the rest to highlight and direct the attention of users in hopes these links represent paths to their final destinations. In order to achieve this goal, best practices found in the literature review were evaluated and molded their approaches to suit the specific needs of this research.

This framework was designed to allow for extensibility; developers should have an easy process of modifying the framework to suit their needs. The framework has methods to return the most popular sites based on the given timeframe, and this timeframe can be altered to provide varying levels of detail. Altering the client-side code is less resource intensive than altering or creating a view in a database, so the framework's method for returning this data resides in the client-side framework. Additionally, developers should be able to modify the user interface with additional metrics, so methods for returning

information based on page views and exit rates are provided. This gives customization options to developers to alter the user interface to match both a given timeframe and provide more detail.

The industry standard jQuery programming language was used to interact with the HTML document object model (DOM). The DOM is a hierarchical, tree structure that lays out the flow of a webpage, including sections, images, text, and links and can be modified with jQuery or other coding languages at load time. By using jQuery, the site's links can be manipulated to apply changes to their styles to increase perceptibility based on their presence in the recommended sites for a given timeframe. Other languages were considered, but jQuery was chosen for its robust feature set which provides additional functionality over JavaScript and is more prevalent in industry as compared to client-side frameworks such as ReactJS or FluxJS.

4.3 Framework Technologies

4.3.1 Cloud Hosting Providers

To create a standalone proxy instance, three major providers of cloud hosting stacks were evaluated based on three criteria: cost, storage, and operation. Google's App Engine, Microsoft's Azure platform, and Amazon's Elastic Cloud Compute (EC2) all provide virtualized instances which can be rented, and upon which developers can create and host their applications. Each of the providers has levels, or tiers, into which their instances

fall, based on storage, bandwidth, and operational requirements. The goal was to find the offering that provided the necessary storage requirements (as it currently stands, the UNF website and all necessary files, images, and code amount to roughly 30GB), appropriate read/write operation limits for the framework, as well as ease of integration and use.

4.3.1.1 Amazon Web Services Technology

Amazon offers a wide variety of solutions in their cloud platform, including solutions for data storage, website hosting, and content delivery, the three most appropriate to this study. Amazon Simple Storage Service (S3) is their robust solution for data storage. It provides 99.999999% reliability and durability. Since the research is conducted with human subjects, security is of the highest priority, and S3 supports encryption both at rest and in transit using SSL encryption (Amazon, 2018). These storage volumes can be tied to VM instances in their Elastic Cloud Compute (EC2) platform or to other areas. AWS also provides website hosting using various content management systems, and content delivery through their Cloudfront content delivery network (CDN) (Amazon, 2108). AWS has wizards available for quickly creating a static website and for creating EC2 VMs that allow clients to quickly create their desired technology stack. Each piece is a pay-as-you-use item and for pennies on the dollar a secure full-stack environment can be set up easily.

4.3.1.2 Google Cloud Technology

Google Cloud offers similar services to AWS in that they provide website and web app hosting, full VM servers, and data storage. Google Cloud offers security with both encryption at rest and encryption in transit for all their services as well as physical security at all their data center locations (Google, 2018). They also provide full operating system VM instances with high availability and a low mean time to recover in the case of data loss. Data is stored on SSD drives as compared to combination of magnetic and solid-state drives with AWS. Google has international data centers, but a far fewer number than AWS maintains across the globe. While not impactful for this research, the closer to a datacenter, the quicker the response for users. Like AWS, Google offer a pay-as-you-go pricing scheme to keep costs to a minimum with alerts and monitoring functionality with transparency into what the technology stack is doing (Google, 2018).

4.3.1.3 Microsoft Azure Technology

Microsoft Azure is Microsoft's cloud computing offering. It includes similar services to both AWS and Google and provides additional functionality out of the box such as Microsoft specific functions like Active Directory and SharePoint directly in the cloud. In addition, they offer full VM instances with ability to host websites, web applications, and data storage. They offer a wide variety of storage options by default, including file, disk, blob or NoSQL storage option. VM instances can range from micro to high-performance instances, but are more limited in the OS options available, favoring

predominantly Windows instances. These VMs can cost slightly more money as they include the licenses for software owned by Microsoft. Without having to stand up VM instance, their integration with Microsoft Visual Studio allows quick remote deployment to Azure Web App services without extensive manual configuration. Their tightly coupled integration with the Microsoft suite of tools does not help us in this research.

4.3.2 Budget

Each of the three instances have varying levels of service for fees as outlined in the Tables 1 to 3 below. Each system offers incentives to use their product: Amazon AWS offers a free tier of service for their instances before you are required to pay for services, Google Cloud compute offers \$300 in credit to use for new accounts. The basic plans are mapped out, not accounting for free options or credits for the three leading systems.

Table 1: Amazon Web Services Pricing

Amazon EC2		
Service	Limit	Price
t1.micro VM	-	\$14.64/month
Analytics API	50,000/day	FREE
Storage	5GB	\$0.50/month
MongoDB	-	FREE
Total	-	\$15.14/month

Table 2: Azure Cloud Services Pricing

Azure		
Service	Limit	Price
A0 VM	-	\$14.88/month
Analytics API	50,000/day	FREE
Storage	5GB	\$0.10/month
	100000 Reads	\$0.50/month
	100000 Writes	\$0.50/month
MongoDB	-	FREE
Total	-	\$15.98/month

Table 3: Google Cloud Services Pricing

Google Cloud Compute		
Service	Limit	Price
AppEngine	-	\$12.19/month
Analytics API	50,000/day	FREE
Storage	5GB	\$0.10/month
	100000 Reads	\$0.50/month
	100000 Writes	\$0.50/month
MongoDB	-	FREE
Total	-	\$13.29/month

There are also additional fees for setting up network access to the servers as well. These are negligible prices that do not affect the overall price much in calculations. Each of the services are roughly the same price per month of use.

4.3.3 Technology Decisions

Amazon AWS was chosen for hosting. With the ease of creating a static website in Amazon Cloudfront with Amazon S3 storage backend, it's possible to cheaply and securely store data. While not the most inexpensive option, prior experience using AWS as a hosting platform and familiarity with the product enabled quick creation of the environment. An AWS EC2 t1.micro instance is being used to utilize the free tier before

having to pay additional fees. A standard Ubuntu Linux VM in AWS was created and on it installed NodeJS and MongoDB in order to create the data store and API. The encrypted storage volume option was used when creating a VM and the MongoDB databases are stored on this drive to increase security of the data. Since the data required is not dependent on Google services except for a single API call once a year, there is no need to create a full system on their resources. A Google Cloud API account is required for querying against the Google Analytics API, but this does not limit use of their software suite. AWS provides all the necessary tools and software to easily integrate the framework solution.

Chapter 5

FRAMEWORK

As part of the research development, the Active Analytics framework was applied on a proxy UNF Website (UNF, 2107). The clone of UNF Website was created in November 2017 as a test instance proxy. Full access to the Google Analytics data for the UNF website was provided with full cooperation and approval from the Information Technology Services at UNF. The test instance was used with a custom active analytics tool to gather experimental data such as success rates and time taken to complete tasks using the proposed system, which is explained further in Chapter 6.

The core focus was implementing changes on the homepage of the UNF website. The homepage contains over 50 clickable items that link to pages within the system (Figure 5). It can be assumed that these represent a view of the most popular categories of information to present to users, according to the university officials. While there is a wealth of information linked to by this page, only those things that should be the most popular based on the historical trends should be highlighted. The goal of this research is to make it easier and take less time for users to be able to find a page they are looking for at a given time of year. In order to achieve this, custom scripts were developed to parse the historical analytics data and apply modifications to the navigation elements on the homepage. The focus of this chapter is on the link selection criteria and the modification used to highlight pages based on those changes.

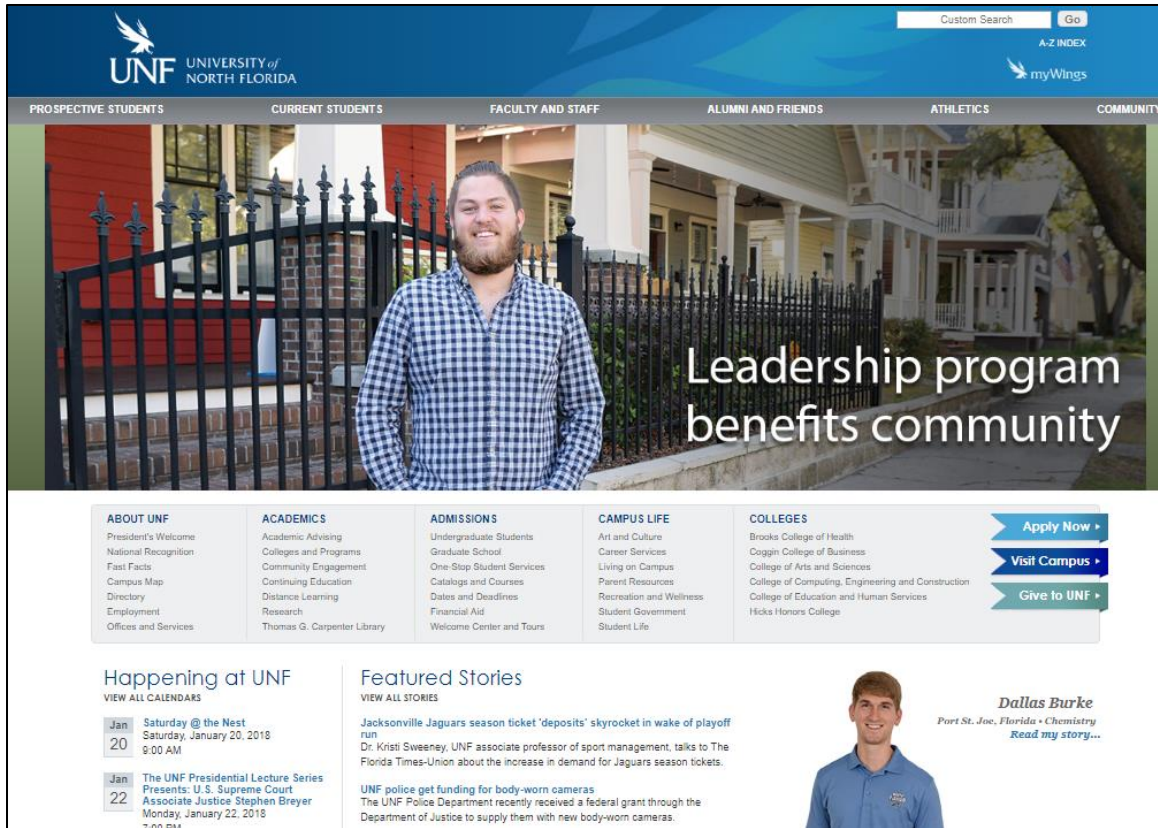


Figure 5: UNF Homepage without Modification

5.1 Link Selection

While developing this framework, statistics were gathered on web page usage for the entire previous year to use for the coming year and break it down by week. Some weeks have over 200 records of popular links (popular links were deemed to be links that were clicked more than ten times in a week) and were sorted from most popular to least popular. Table 4 shows the top 10 pages for three weeks in 2016.

Table 4: Top Webpages for Weeks 5-7, 2016

Week 05	Week 06	Week 07
/default.aspx	/default.aspx	/default.aspx
/library/default.aspx	/library/default.aspx	/library/default.aspx
/admissions/default.aspx	/admissions/default.aspx	/admissions/default.aspx
/prospects/default.aspx	/prospects/default.aspx	/prospects/default.aspx
/info/academics/default.aspx	/catalog/programs/ug/default.aspx	/catalog/programs/ug/default.aspx
/catalog/programs/ug/default.aspx	/info/academics/default.aspx	/info/academics/default.aspx
/graduateschool/academics/Graduate_Programs.aspx	/current/default.aspx	/current/default.aspx
/current/default.aspx	/graduateschool/academics/Graduate_Programs.aspx	/catalog/academic-calendar/academic_calendar/default.aspx
/housing/default.aspx	/housing/default.aspx	/graduateschool/academics/Graduate_Programs.aspx
/tuition/default.aspx	/admissions/applynow/default.aspx	/catalog/default.aspx

As shown in the table above, frequently, the same sites appear in the top ten of each week, at different rankings each week. The top two, the default or homepage of UNF and the Library page were the top two pages in every week. These were removed from the selection for two reasons. The first, the default page is going to be the most popular page as it is the typical landing page, and since it is the page being modified there is no need to select this page. Secondly, the Library page is also another landing page; any user actions taken on the library page redirects to another CMS system that is separate from the main UNF website. Since the Library CMS is not part of the UNF site, it is not possible to duplicate and apply the tracking scripts to it. Therefore, these pages are not considered when selecting links to highlight.

To query the data from the analytics store, an API endpoint was created. The first step taken is to calculate the current week of the year, represented as a numerical value from 1

to 52. These weeks follow the standard week where Sunday is the first day and Saturday is the last. The current week number is passed to the API endpoint and from there the numbers of the previous and next weeks are calculated. For example, consider January 25th, 2018 date as the user visit date, which would be week 3 in 2018. The data store is queried and documents for weeks 2, 3, and 4 from the year 2017 are selected. These documents received from each week query are aggregated and ordered by overall page views. The default page and the library page are removed for the reasons discussed above along with duplicate values, extraneous items such as 404 error pages, and the resultant set is returned as a JSON object.

Figure 6 below shows a selection of code which was used to query the data. While the API call returns all the links from the previous, current, and next weeks after the processing explained above, the interaction is limited to the top 10 links for processing the UI modifications.

```
function getWeekData(week) {
    var weeks = getWeeks();

    apiUrl = "http://api.jacobkoza.com:8080/API/GetSlidingWindow/" + String(weeks)

    $.ajax({url: apiUrl,
        success: function(data){
            var top5 = data.slice(0,5)
            var secondary = data.slice(5,10)
            for(var i = 0; i < top5.length; i++){
                updateLinks(top5[i], "top")
            }
            for(var i = 0; i < top5.length; i++){
                updateLinks(secondary[i], "secondary")
            }
        }
    });
}
```

Figure 6: API Call to Get Top Links

There has not been much research into the optimal number of links to highlight on a given page, but the selection of 10 items was not fully arbitrary. Research conducted by Sheena Iyengar in her book “The Art of Choosing” shows that too many options can be debilitating to people (2011). In her research, focused on selling jams, she found that while presenting more options to people enticed them to view the display more often, people bought jam more often when their selection was limited (Iyengar, 2011). The allusion of choice may be nice, but also quickly may become overwhelming. Along these research lines, highlighting more links would be a detriment as when everything becomes highlighted, nothing stands out. Also, based on analysis, almost all weeks have at least one difference in top ten sites compared to its surrounding weeks, therefore, the visual modifications will be changed depending on the week a user accesses the site.

5.2 Homepage Modifications

Two subtle different modification styles are applied depending on the week that the specified link came from analytics data. For the top five most popular links for the current week last year, a styling of a dark blue background and bold font was chosen to highlight its importance. For the links that were returned from the query for the next second five, a slightly lighter blue color than the top five but different enough from the standard background to distinguish it was chosen. The bold font and the color were chosen for the top links so that it draws the user’s eye quicker to things they may need or want. Figure 7 below shows an example of how the custom styling based on the week selector is applied.

```
function updateLinks(element, selector){
  var url = element
  if(selector == "secondary"){
    |   $('a[href^="'+url+'"]').css({"color": "white", "background-color": "#71c7ec", "font-weight": "light"});
  }else if(selector == "top"){
    |   $('a[href^="'+url+'"]').css({"color": "white", "background-color": "#189ad3", "font-weight": "bold"});
  }else{}
}
```

Figure 7: CSS modification code for popular links

In addition to modifying the hyperlink portions, a few other modifications were made. Firstly, the ability to use the custom search box across the site was removed; this will prevent users from searching for the page and skewing the results. Users were restricted to only navigating through the navigational items on the pages themselves. Secondly, users were restricted from clicking on links that would take them to external sites where they could not be tracked since this data would have no bearing on this study. Other research has gone so far as to create sections on pages which display the most popular links across the entirety of the site, but this would cause unneeded complexity. Research has shown the most common landing page for the test site is the homepage, and the goal is to provide visual indications of logical next links on that page alone. An example of the site with its visual modifications can be seen below in Figure 8.

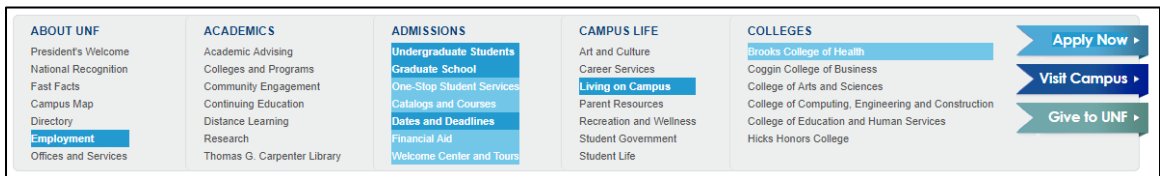


Figure 8: Select UNF Links with Proposed Modifications

Chapter 6

FRAMEWORK EVALUATION

As mentioned previously, two alternate versions of the UNF homepage were tested, one original version with no modifications, and one with the adaptive analytics framework applied. Regarding evaluation measures, the focus was on time taken to complete a set of navigation tasks, number of clicks taken to reach a page, success rates of said tasks, and overall usability aspects of the web site on either version. This chapter expounds the test plan for these metrics.

6.1 Evaluation Goals

For this research, the goal was to see if there is significant difference between the original UNF site, and the site with the proposed framework. To accomplish this, randomized A/B testing was used, in which visitors to the site were randomly assigned either the original site or the modified version. The results of the study were then compared for each site to determine statistical differences. The testing process is explained in section 6.2.

Participants were assigned a set of tasks to be completed (see Appendix A for full task details) while their actions were being automatically recorded. The focus of the testing was on a few main variables including the amount of time it takes from beginning a task to reaching the desired page, and the number of clicks taken to find the desired page.

Following the navigation tasks, volunteers will be presented with a short survey to gather

a few demographic data points in addition to asking questions about their feeling of the usability of the site. Appendix B lists the full survey.

Additional variables that were outside of control were identified. The choice of population could have the potential to skew the results. Targeting members of the UNF community has a high probability that the volunteers have already interacted with the UNF site. Having used the site previously an individual may have prior knowledge of the location of a page and how to get there. To mitigate against skewed results, the target audience included both faculty and staff to attempt get varying levels of interaction with the UNF page as well as technical acuity. This should provide a diverse enough population to prevent data skew.

6.2 Testing Process

The testing process will be a standard A/B testing of the site (Salkind, 2010). A/B testing is a standard way to evaluate two different version of the same site. A/B testing is an industry standard way of evaluating changes in two version of the same website. Web designers typically use A/B testing to determine if interface changes, such as moving a button, changing wording, or altering a color scheme has any impact on conversion rates, bounce rates, and/or user satisfaction. Tracking allows designers to choose which version of the UX is more successful by their target metrics and can adopt changes accordingly.

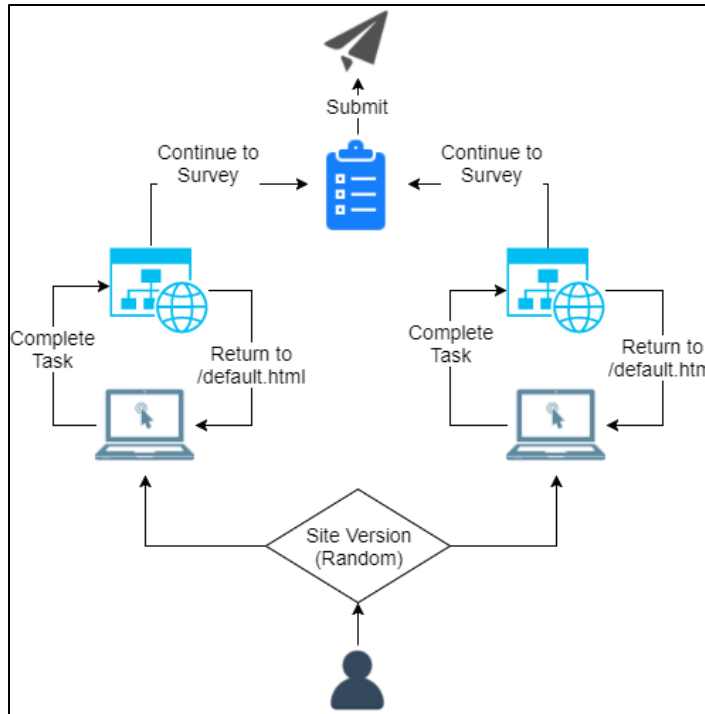


Figure 9: Framework Testing Procedure

For this research, a slightly less straightforward approach was taken. As the UNF website is a live system used by the school for marketing purposes, it was not possible apply the framework directly to the site. This limits the amount of data available as there is no access to the actual servers. Instead, a cloned version of the website was created and had the framework applied to it. In this scenario, two versions of the UNF site were randomly presented, one with the framework enabled, and one without. On both versions, the same eight tasks were presented to users, including tasks such as “navigate to the course catalog page.” See Appendix A for a full list of task prompts. The control group for this study were the users assigned to the standard UNF site. They were asked to navigate without the aid of the framework. The treatment group was the the group assigned to the framework-enabled version of the site. Following the tasks, users were asked to fill out a

short survey answering questions about their technological acuity and the overall impressions of the site (see Appendix B). Tracking the completion metrics for a larger number of tasks should help offset the relatively small sample size.

To test the goals, as mentioned previously, eight pages of the UNF website were identified that users were directed toward (see Appendix A for the full list). These links were selected semi-randomly to avoid any potential bias. In order to achieve the semi-random selection of destinations, pages were selected based on their perceived importance to both students and staff. Section 6.3 below discusses the participants targeted for the research, but both students and staff were included for evaluation. Tasks were selected that would pertain to both groups of people; six tasks were directed at students or potential students including pages such as tuition and fees, and two were employment pages geared toward staff members, such as employee benefits.

6.3 Participants

For this research, members of the UNF community were recruited as communication with them was easiest. Each participant needed to voluntarily engage in the study and had permission to quit or refuse at any time without penalty. The study took only roughly 15 minutes of their time and no compensation was offered for those who took the test. The study was kept as short as possible to minimize the number of volunteers who quit mid-test and increase the number of completions. To that end, tasks that were either to common pages or held pertinent information for anyone using were considered as well as

choosing tasks where few steps were needed to reach the desired page. An automated testing system was used to track when a user would click on a hyperlink and record each click in conjunction with its timestamp. Three recruitment strategies were used for this, two for students, and one for faculty and staff. Requests were sent to the UNF School of Computing faculty requesting they include the information on their Canvas sites.

Information was also sent out campus-wide communications through the Osprey Update for both staff and students requesting volunteers for this study. Each instance provided information about the study and provided the link to access the study.

6.4 Institutional Review Board (IRB) Approval

Since this research involved the use of human subject, approval of the Institutional Review Board for UNF was required before conducting the research. This research description and all accompanying documentation was submitted for review on January 16, 2018. Approval was granted by IRB on March 8, 2018. See Appendix C for full IRB approval letter.

Chapter 7

EVALUATION RESULTS AND ANALYSIS

For this research, qualitative and quantitative data was collected to attempt to understand a big picture view of the effect of the framework on the specified goals. The focus was on the following quantitative measures as key indicators of the success or failure of the hypothesis: number of steps taken to complete a given task and the amount of time taken from task initiation to completion. The solution was tested using A/B testing (Salkind, 2010), where one group of users, randomly selected, was assigned a version of the UNF website with the framework enabled, and the other was assigned the plain site without modification. Using t-Test analysis the difference between the two sample groups was evaluated. The t-Test results for both completion time and navigation steps are explained in section 7.2.

In addition, qualitative data was also collected to see if patterns emerged between the data and the results of the quantitative analysis: demographic information about the users age, use of the Internet, and familiarity with the UNF site, a mean rating of the overall feeling of users towards the site, and free form response answers. The few free-form answers provided did not garner enough responses to make any meaningful analysis as they were optional fields that did not require completion. The focus of these qualitative responses was to further strengthen the hypothesis by looking at the emotional response to the framework site compared to the original to prove the framework-enabled version

made navigation easier and, overall, more enjoyable. The demographic details are evaluated in section 7.1, followed by the quantitative analysis.

7.1 Demographics

Following the completion of all eight tasks, the survey volunteers were asked to provide demographic information about themselves. This data included age, experience with the Internet, and the frequency visits to the UNF site. The average demographic age of UNF students was between 18 and 25 and to attempt to gather a larger range of population, faculty and staff were included. Another set of goals was also to try to involve individuals with a variety of experience in using the Internet, and the UNF web site. Figures 10-14 below show the demographic breakdown of users who took the survey. Each graph shows all possible responses and the frequency of users who fall into each of the categories. As shown by the data, the research participants skewed towards a younger, more internet experienced population with the 18-25 and 26-35 population groups being the largest. This is congruent with the demographic of the University, with over 16000 students and roughly 4000 employees. The lack of users with limited Internet experience can likely be attributed to the growth of the Internet as a daily interaction for people.

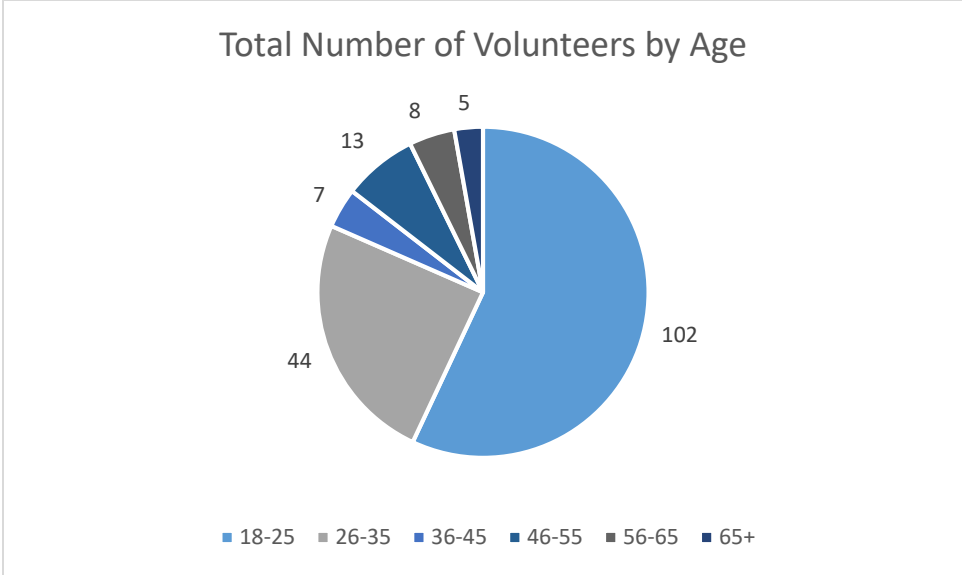


Figure 10: Responses by Age

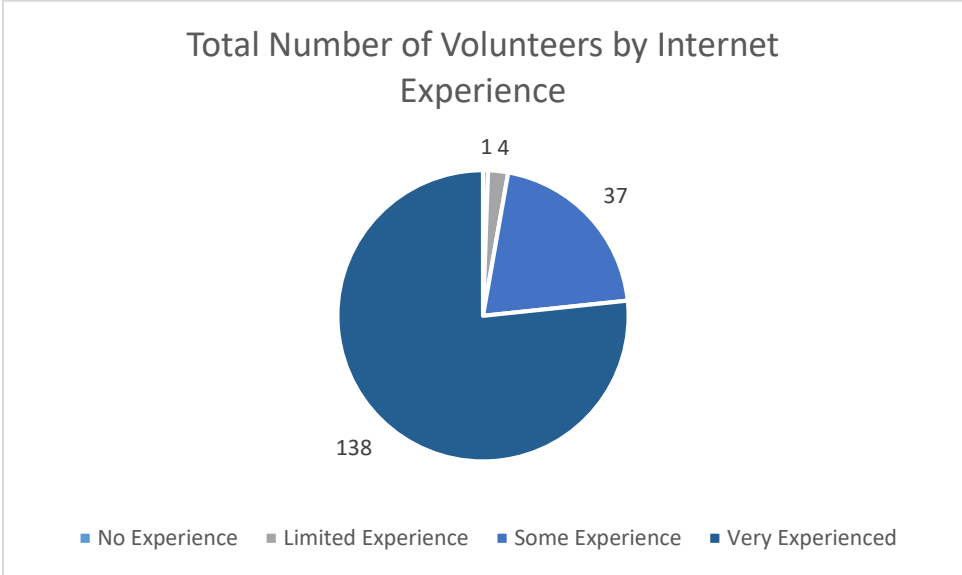


Figure 11: Responses by Internet Experience

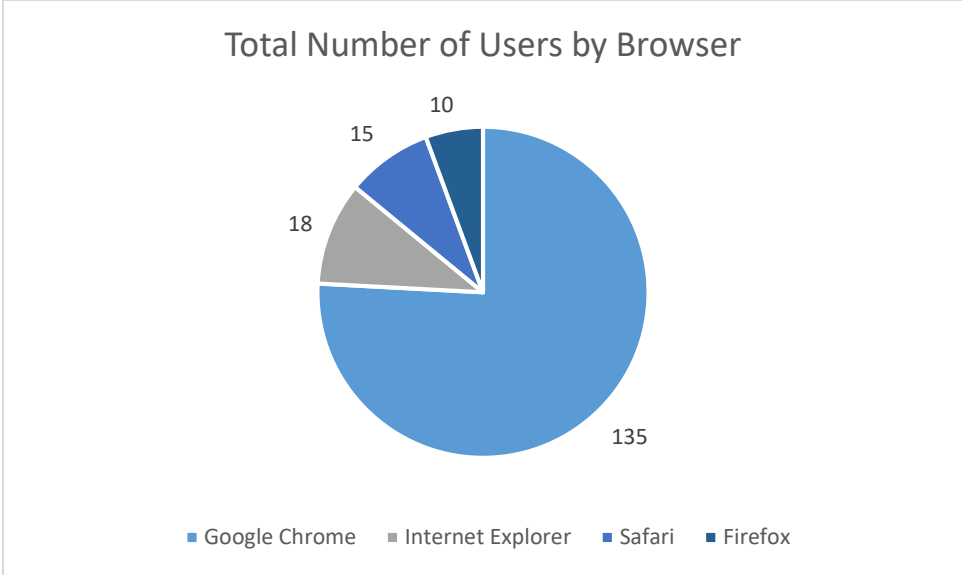


Figure 12: Reponses by Browser

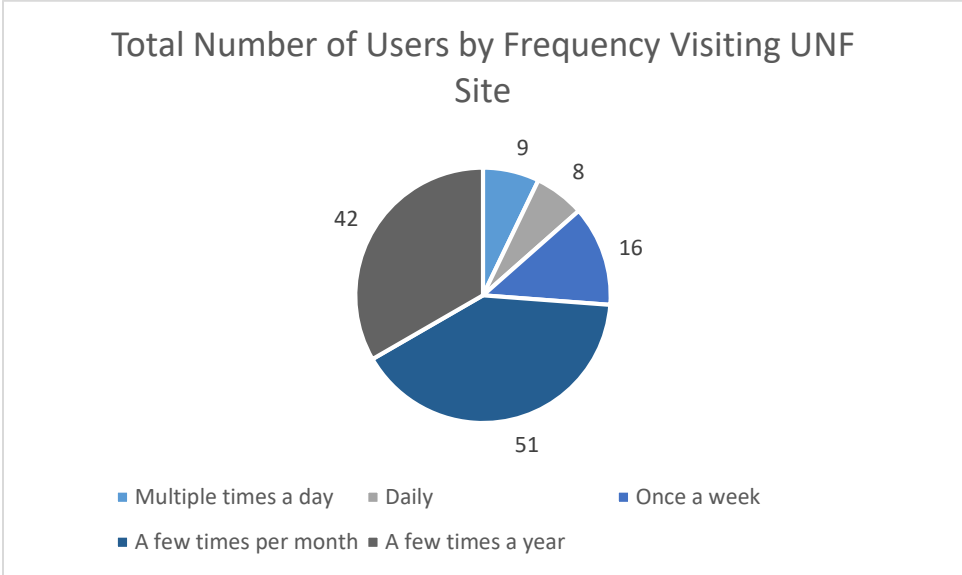


Figure 13: Reponses by How Often Users Visit unf.edu

7.2 Independent Samples t-Test Data

For each of the quantitative metrics, the two response groups were compared and evaluated to see if the difference was statistically significant or not to reject the null hypothesis, or if there was insufficient data, failing to reject the null hypothesis using independent t-Tests (Salkind, 2010). For these tests to be considered valid, three assumptions must be met:

- Assumption 1: The data must be observed for the two groups independently, and every observation cannot predict the observation of another in the test. In this research, each user was only able to take the survey once, ensured by providing their identity. There was no dependency between individuals to complete the task; each user took the survey alone and the responses had no predictive ability between users.
- Assumption 2: The second assumption involves the equality of variance in the sample groups. Levene's test for equal variance is used to determine whether one could assume there was equal variance between the two testing groups, represented by a p value of > 0.05 . In this analysis, if the population met this criterion, the t-Test results assumed equality of variance, and if not, the t-Test results without assuming equal variance were evaluated. Both values are represented in this analysis.
- Assumption 3: The final assumption is that the observation samples must be pulled from a population that follows a normal standard distribution. Each of the measures used was tested using the Shapiro-Wilk test (Salkind, 2010) to

determine if the values follow a normal distribution. If any value did not follow a normal distribution, the data was transformed to fit.

7.2.1 Task Completion Time

For each task assigned to the volunteers, the time taken to complete the task was measured by recording the time each link was clicked. For this metric, the hypothesis was that users of the framework-enabled version should take less time to reach the desired site as compared to the plain version. The null hypothesis for this measurement would be that it would take the same amount of time to reach the end page whether the framework was enabled or not. Figure 14 shows the average time it took users to complete each of the eight tasks assigned to them, both using the framework-enabled site as well as the non-enabled site.

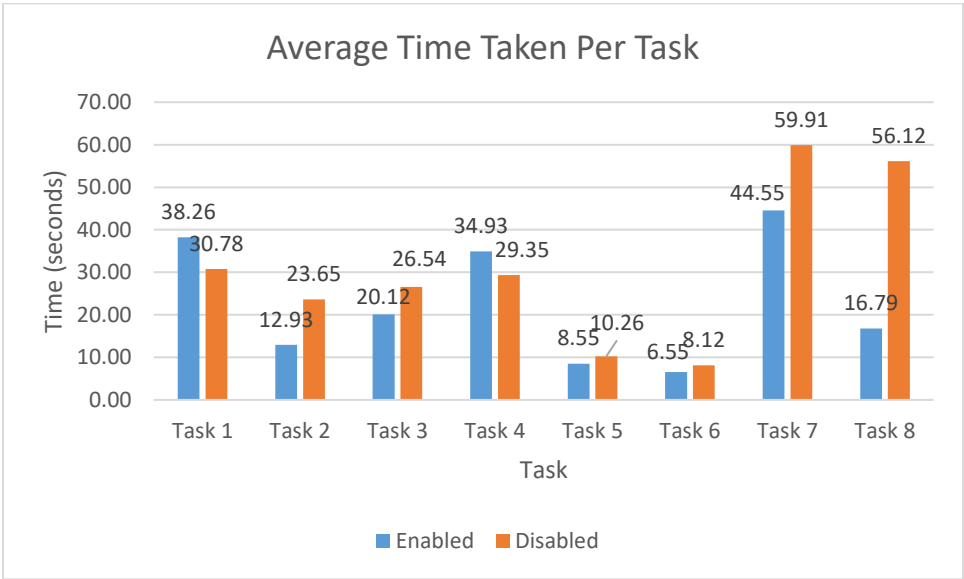


Figure 14: Time Taken to Complete Each Task

As you can see from the chart above, six of the eight tasks assigned to users of the framework-enabled site were able to be completed in less time than those with the framework not enabled. Tasks 2, 3, 4, 5, 6, and 7 were completed in less time, and the remaining tasks were completed in greater time. For each of the eight tasks, independent sample t-Tests were conducted to determine if the differences were statistically significant or not.

7.2.1.1 Time Taken t-Test Results

For the number of responses, the data for time taken to complete each task was not normalized, as validated using the Shapiro-Wilks test for normality. This means that none of the tasks, either with the framework enabled or framework disabled fit a normal distribution. Each of the tasks had a positive skew, meaning there were more values left of the bell curve than to the right. To try to get the data to fit a more normal distribution, a logarithmic transformation was applied. Taking the log base 10 of the time values is a valid way to attempt to normalize positively skewed data. Table 5 below shows the statistical values of the data before applying the log transformation to each of the tasks. Table 6 shows the values of each tasks after applying a log base 10 transformation to the data.

Table 5: Time Taken Statistics - Non-Normalized

	Framework	N	Mean	Std. Dev	Std. Err Mean	Skewness	Kurtosis
Task 1	Enabled	84	50.10714	97.59148	10.6481	3.494664	15.31464
	Disabled	65	36.93846	50.25618	6.233512	1.787962	2.109555
Task 2	Enabled	86	16.53488	26.66114	2.874945	2.97965	9.919604
	Disabled	66	27.95455	40.4815	4.982928	2.934536	10.59382
Task 3	Enabled	84	26.34524	35.19395	3.839975	2.450976	6.056857
	Disabled	67	30.89552	38.66959	4.724243	2.410821	5.45909
Task 4	Enabled	83	46.28916	46.23628	5.075091	1.769334	2.993488
	Disabled	65	35.21538	30.58211	3.793244	2.050641	4.665408
Task 5	Enabled	84	11.19048	26.96069	2.941653	6.129611	43.51321
	Disabled	65	12.30769	15.99504	1.983941	2.644598	9.552382
Task 6	Enabled	85	8.470588	12.57699	1.364166	3.591963	13.71081
	Disabled	66	9.590909	16.4251	2.02179	4.286036	21.24485
Task 7	Enabled	85	57.64706	113.6703	12.32928	4.636088	25.5444
	Disabled	66	70.80303	81.12065	9.985262	2.898533	11.41688
Task 8	Enabled	80	23.0875	43.37041	4.848959	5.441122	36.43046
	Disabled	65	67.33846	363.9516	45.14264	7.917493	63.36817

Table 6: Time Taken Statistical Values

	Framework	N	Mean	Std. Dev	Std. Err Mean	Skewness	Kurtosis
Task 1	Enabled	84	1.163302	0.653963	0.071353	0.707917	-0.59962
	Disabled	65	1.169821	0.61792	0.076643	0.214902	-1.16428
Task 2	Enabled	84	0.850648	0.469495	0.051226	0.735545	-0.32651
	Disabled	66	1.159918	0.432424	0.072094	0.238064	2.801198
Task 3	Enabled	84	1.159918	0.450663	0.049171	0.659284	-0.47142
	Disabled	67	1.259858	0.432424	0.052829	0.424099	-0.19562
Task 4	Enabled	83	1.48007	0.404256	0.044373	0.18583	-0.64309
	Disabled	65	1.420409	0.32771	0.040647	0.238064	-0.15761
Task 5	Enabled	83	0.727709	0.417927	0.045873	1.675522	2.801198
	Disabled	65	0.794472	0.510615	0.063334	0.344849	-1.00502
Task 6	Enabled	85	0.714956	0.375698	0.04075	1.073141	1.223822
	Disabled	66	0.73079	0.405821	0.049953	1.013433	1.377908
Task 7	Enabled	85	1.395594	0.508113	0.055113	0.825761	0.113991
	Disabled	66	1.625483	0.468981	0.057728	-0.30475	0.14397
Task 8	Enabled	80	1.049637	0.492043	0.055012	0.400898	0.128789
	Disabled	65	1.068428	0.571447	0.070879	1.400583	3.925381

As can be seen from the two tables above, the skewness and kurtosis were reduced to closer to normal levels. Kurtosis with a high value indicates data that has heavy tails, meaning there are many outliers in the data; low levels of kurtosis indicate a lower level of outliers. Levels between -2 and +2 are considered acceptable to prove normal distribution (George & Mallery, 2010). Positive and negative skew indicate the predicate of data to the left and the right of the median, respectively. In perfectly normal data, skewness would be 0, but ranges between -1 and +1 indicate low to moderate skew in a normal distribution (George & Mallery, 2010).

Based on the results of the Shapiro-Wilk Test, five of the 16 tasks (eight with the framework enabled and eight without) statistically fit a normal distribution. While t-Test analysis of small populations that do not fit a normal distribution does not have any predictive power, populations of greater than 30 can be assumed to have normal distribution (Salkind, 2010). All eight tasks, from both versions, had population sizes of greater than 60. Any further modification to the data would reduce its predictive nature. The normalized data was then taken independent sample t-Tests were conducted on it to determine if there was any statistical difference between the time taken to complete a task with the framework enabled or disabled. Table 6 below shows the values of those calculations.

Table 7: Time Taken t-Test Results

	Leven's Test	t-Test for Equality of Means				
	Sig	t	df	Sig (2 tail)	95% Confidence Interval	
					Lower	Upper
Task 1						
Eq. Var. Assumed	0.82466713	0.061799	147	0.950807	-0.21497	0.201935
Eq. Var. Not. Assumed		0.062249	141.2178	0.950452	-0.21353	0.200496
Task 2						
Eq. Var. Assumed	0.00422558	2.668745	148	0.008463	-0.40013	-0.05967
Eq. Var. Not. Assumed		2.599471	122.7069	0.010482	-0.40497	-0.05483
Task 3						
Eq. Var. Assumed	0.35974747	1.378293	149	0.170179	-0.24322	0.043341
Eq. Var. Not. Assumed		1.384758	143.9688	0.168269	-0.24259	0.042713
Task 4						
Eq. Var. Assumed	0.0301324	0.966639	146	0.335323	-0.06232	0.181641
Eq. Var. Not. Assumed		0.991442	145.81	0.323112	-0.05927	0.178591
Task 5						
Eq. Var. Assumed	0.00261961	0.874639	146	0.383207	-0.21762	0.084095
Eq. Var. Not. Assumed		0.853715	122.4604	0.39493	-0.22157	0.088041
Task 6						
Eq. Var. Assumed	0.6444988	0.248033	149	0.80445	-0.14198	0.110315
Eq. Var. Not. Assumed		0.245626	134.2812	0.806346	-0.14334	0.111666
Task 7						
Eq. Var. Assumed	0.24513998	2.851373	149	0.004971	-0.3892	-0.07058
Eq. Var. Not. Assumed		2.880404	144.5591	0.004576	-0.38764	-0.07214
Task 8						
Eq. Var. Assumed	0.50140773	0.212705	143	0.83186	-0.19342	0.15584
Eq. Var. Not. Assumed		0.209439	126.9962	0.834441	-0.19634	0.158754

As can be seen from the table above, the null hypothesis can be rejected on two of the tasks for this time-taken metric. Tasks 2 and 7 both were statistically shorter when the framework was enabled. Levene's test for Equality of Variances was conducted for all tasks, and each of the eight tasks has a significant difference between the two versions ($p > 0.05$). Each of the eight tasks are analyzed below.

Task 1: Colleges and Programs

For the first task, users were asked to navigate to the page showing the description of catalogs and programs for the University. According to Figure 14, users were able to reach the page faster without the framework enabled. Levene's test for Equal Variances show equal variances can be assumed, so the value for that record in Table 6 is taken. Based on the results of the t-Test analysis, the p value being greater than 0.05 indicates there is not a significant difference between the two. For this task, the null hypothesis cannot be rejected as the difference is not statistically significant.

Task 2. Course Catalog

For this task, users were asked to navigate to the page showing the listing of the course catalog with class descriptions. With the framework enabled, users were able to navigate to the page almost 11 seconds quicker with the framework than without. Levene's test was significant for this task, so the value from the unequal variances assumed category is taken. Looking at the Sig value for the t-Test analysis, it does indicate a statistically

significant change ($p < 0.05$) for the 95% confidence interval. For this task, the null hypothesis can be rejected.

Task 3. Graduate School Programs

For this task, the users were asked to navigate to the page showing the graduate school programs available. With the framework enabled, users were able to navigate to the page an average of 6 seconds faster with the framework. According to Levene's test ($p > 0.05$) equality of variance can be assumed between the two groups, so the value from the equal variances assumed category is used. Looking at the Sig value for the t-Test analysis, it does not indicate a statistically significant change ($p > 0.05$) for the 95% confidence interval. For this task, the null hypothesis cannot be rejected as the difference is not statistically significant in favor of the framework.

Task 4. Tuition

For this task, the users were asked to navigate to the page showing the list of tuition and fees for the university. With the framework enabled, took an average of 6 second less to reach the desired page when the framework was disabled. Users were able to find the page quicker in the original site. Levene's test was significant for the two test groups so the value from the unequal variances assumed category is taken. Looking at the result of the t-Test analysis, there is no statistically significant change ($p > 0.05$) between the two groups for the 95% confidence interval. For this task the null hypothesis cannot be rejected.

Task 5. Admissions Deadlines

For this task, users were asked to navigate to the page showing the listing of the deadlines for applying to the university. With the framework enabled, users were able to navigate to the page almost 2 second quicker when the framework was enabled. Levene's test was not significant for this task, the value from the equal variances assumed category in Table 6 above was used. Evaluating the results of the t-Test analysis, it does not indicate a statistically significant change ($p > 0.05$) for the 95% confidence interval. For this task, the null hypothesis cannot be rejected as the difference is not statistically significant.

Task 6. Apply

For this task, the users were asked to navigate to the page to begin an application with the university. With the framework enabled, users were able to navigate to the page an average of 1.5 seconds faster when the framework was enabled as compare to not. Levene's test was not significant for this task, the value from the equal variances assumed category above is selected. Looking at the Sig value for the t-Test analysis, it does not indicate a statistically significant change ($p < 0.05$) for the 95% confidence interval. For this task, the null hypothesis cannot be rejected as the difference is not statistically significant.

Task 7. HR Student Employment

For this task, the focused changed from trying to access student resources to accessing faculty and staff resources, specifically the student employment page. With the framework enabled, users took an average of 15 seconds less time to reach the desired page when the framework was enabled. Levene's test was not significant for this task, so the value from the equal variances assumed category was taken. Looking at the Sig value for the t-Test analysis in Table 6 above, it does indicate a statistically significant change ($p < 0.05$) for the 95% confidence interval. For this task, the null hypothesis can be rejected as the difference is statistically significant.

Tsk 8. HR Benefits

For this task, the users were asked to navigate to the page that explained the benefits available to employees of the university. With the framework enabled, users were able to navigate to the page an average of 40 seconds quicker when the framework was enabled. Levene's test was not significant for this task, so the value from the equal variances assumed category is taken. Looking at the Sig value for the t-Test analysis, it does not indicate a statistically significant change ($p > 0.05$) for the 95% confidence interval. For this task, the null hypothesis cannot be rejected as the difference is statistically significant.

7.2.2.2 Time Taken Summary

The time taken was statistically improved for two of the eight tasks presented to users, but some tasks took longer to complete with the framework. Specifically, Task 1, was a task that users took longer with the framework. During the testing period, the link on the homepage for this task for the shorted distance was not highlighted. In the preliminary research, in some weeks this link would have been highlighted. One possible explanation is that the presence of highlighted links that did not lead to the desired end were a distraction. Other tasks, such as 2 and 7, the statistically significant changes did have their links highlighted on the home page. Task 8 was also slightly distorted in its mean difference of time due to one response being more than 10 minutes, far larger than the rest of the responses. This proves the need for log transformation to normalize the data, even if this task was not statistically better with the framework.

7.2.3 Navigation Steps Taken

In addition to tracking the amount of time taken to complete each task, the number of clicks taken to traverse from homepage to desired page was taken. The hypothesis was that when the framework was applied, the average number of steps taken would be less than when users interacted with the plain site. The null hypothesis is that the number of steps taken would be the same between the two versions of the site. Figure 15 below shows the mean of steps taken for each task for both the framework-enabled and non-enabled version of the site.

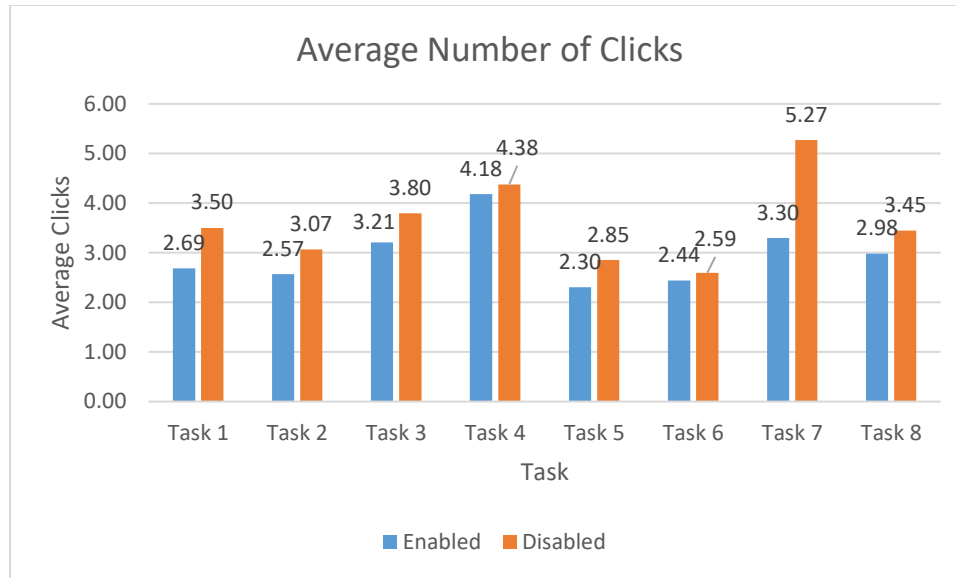


Figure 15: Steps Taken to Reach Desired Page

The graph above shows the average number of steps taken grouped by tasks for both the framework-enabled and non-enabled sites for testing. Based on research, illustrated in Figure 15 above, all eight tasks were completed in less steps when the framework was enabled. While the average shows that all tasks were completed in less steps when the framework was enabled, independent t-Test analysis was conducted to determine if the difference in these values was statistically significant enough to reject the null hypothesis mentioned previously.

7.2.3.1 Navigation t-Test Results

When measuring the statistics for this metrics, the values tended to skew toward lower number of steps taken, meaning less than 4 steps. The data for these tasks does not follow

the standard normal distribution, but because the values are small, normalizing the data would have a nominal effect for analysis. While assumption three above requires that data follow a normal distribution, data sets of large enough sample sizes, where N is greater than 30, data does not have to strictly adhere to a normal distribution (Salkind, 2010). Table 7 below shows the mean, standard deviation, and standard error for the number of navigation steps for each task. Table 7 below shows the results of the independent sample t-Tests and Levene's test for the assumption of equal variances. Using both tables validates whether the number of navigation steps taken for each task with the framework enabled was able to reject the null hypothesis, or if the data fails to reject the null hypothesis for any of the tasks.

Table 8 Navigation Step Data

	Framework	N	Mean	Std. Dev	Std. Err Mean	Skewness	Kurtosis
Task 1	Enabled	90	3.166667	2.204438831	0.232368	3.934998	20.54194
	Disabled	63	4.111111	2.979018626	0.375321	1.729903	3.432881
Task 2	Enabled	100	2.7	1.500841515	0.150084	2.59578	6.891541
	Disabled	68	3.338235	1.66267447	0.201629	1.08538	0.573265
Task 3	Enabled	96	3.510417	1.305815266	0.133274	2.801193	7.510748
	Disabled	66	4.257576	1.986962166	0.244578	1.669697	1.978733
Task 4	Enabled	91	4.824176	2.681597395	0.281108	1.864003	2.957224
	Disabled	67	4.835821	2.626191014	0.32084	2.698442	8.911673
Task 5	Enabled	100	2.42	0.923104118	0.232368	2.598389	7.555599
	Disabled	71	2.971831	1.423996699	0.232368	1.455947	1.612052
Task 6	Enabled	102	2.509804	0.972341239	0.232368	3.827828	20.60191
	Disabled	71	2.704225	1.126237303	0.232368	3.453594	15.7987
Task 7	Enabled	86	4.023256	2.222752635	0.232368	2.760313	8.344769
	Disabled	61	6.393443	4.779395668	0.232368	2.639119	8.61429
Task 8	Enabled	87	3.597701	1.315911304	0.232368	3.10454	10.31743
	Disabled	68	3.75	1.124689012	0.232368	1.682179	2.212718

Table 9: Levene's Test and t-Test Results

	Leven's Test	t-Test for Equality of Means				
	Sig	t	df	Sig (2 tail)	95% Confidence Interval	
					Lower	Upper
Task 1						
Eq. Var. Assumed	0.002586017	2.253696	151	0.025654997	-1.77243	-0.11646
Eq. Var. Not. Assumed		2.139509	107.624	0.034654438	-1.81947	-0.06942
Task 2						
Eq. Var. Assumed	0.020036627	2.589329	166	0.010470374	-1.12489	-0.15158
Eq. Var. Not. Assumed		2.539176	133.9781	0.01225415	-1.13537	-0.1411
Task 3						
Eq. Var. Assumed	0.000146653	2.888803	160	0.004403238	-1.25795	-0.23637
Eq. Var. Not. Assumed		2.682482	103.1118	0.00851285	-1.29956	-0.19476
Task 4						
Eq. Var. Assumed	0.526227984	0.027212	156	0.978325039	-0.85693	0.833642
Eq. Var. Not. Assumed		0.027299	143.9963	0.978258684	-0.85479	0.831498
Task 5						
Eq. Var. Assumed	1.90517E-05	3.072803	169	0.002472594	-0.90635	-0.19731
Eq. Var. Not. Assumed		2.865686	111.0143	0.00497854	-0.93341	-0.17025
Task 6						
Eq. Var. Assumed	0.61448017	1.211743	171	0.227282477	-0.51113	0.122292
Eq. Var. Not. Assumed		1.180285	136.091	0.239945657	-0.52017	0.131329
Task 7						
Eq. Var. Assumed	0.000157154	4.029337	145	8.99507E-05	-3.5328	-1.20757
Eq. Var. Not. Assumed		3.606465	78.51749	0.000543607	-3.67845	-1.06193
Task 8						
Eq. Var. Assumed	0.862380691	0.761359	153	0.447614542	-0.54749	0.242889
Eq. Var. Not. Assumed		0.776131	151.7413	0.438879765	-0.53999	0.235394

Task 1: Colleges and Programs

For the first task, users were asked to navigate to the page showing the description of catalogs and programs for the university. With the framework enabled, users were able to navigate to the page almost a full click fewer when the framework was enabled. Levene's test was significant, so the value from the unequal variances assumed category is taken. Looking at the Sig value for the t-Test analysis, it does indicate a statistically significant change ($p < 0.05$). For this task, the null hypothesis can be rejected for when the framework is enabled.

Task 2. Course Catalog

For this task, users were asked to navigate to the page showing the listing of the course catalog with class descriptions. With the framework enabled, users were able to navigate to the page an average of 0.5 clicks fewer with the framework enabled. Levene's test was significant for this task, so the value from the unequal variances assumed category is taken. Looking at the Sig value for the t-Test analysis, it does indicate a statistically significant change ($p < 0.05$) for the 95% confidence interval. For this task, the null hypothesis can be rejected as the difference is statistically significant.

Task 3. Graduate School Programs

For this task, users were asked to navigate to the page showing the graduate school programs available. With the framework enabled, users were able to navigate to the page

an average of 0.6 clicks when using the framework enabled site. Levene's test was significant for this task ($p < 0.05$), so the value from the unequal variances assumed category is used. Looking at the Sig value for the t-Test analysis, it does indicate a statistically significant change ($p < 0.05$) for the 95% confidence interval. For this task, the null hypothesis can be rejected as the difference is statistically significant that the framework was better for the users.

Task 4. Tuition

For this task, users were asked to navigate to the page showing the list of tuition and fees for the university. With the framework enabled, it took an average of 0.2 clicks fewer to reach the desired page than when the framework was not enabled. Levene's test was not significant ($p > 0.05$) for this task, so the value from the unequal variances assumed category is used. Evaluating the results of the t-Test analysis, it does not indicate a statistically significant change ($p > 0.05$) for the 95% confidence interval. For this task, the null hypothesis cannot be rejected as the difference is not statistically significant.

Task 5. Admissions Deadlines

For this task, users were asked to navigate to the page showing the listing of the deadlines for applying to the university. With the framework enabled, users were able to navigate to the page an average of 0.5 clicks fewer when the framework is enabled. Levene's test was significant for this task, so the value from the unequal variances assumed category is used. Looking at the Sig value for the t-Test analysis, it does indicate a statistically

significant change ($p < 0.05$) for the 95% confidence interval. For this task, the null hypothesis can be rejected as the difference is statistically significant.

Task 6. Apply

For this task, users were asked to navigate to the page to begin an application with the university. With the framework enabled, users were able to navigate to the page an average of 0.1 clicks fewer than with the framework not enabled. Levene's test was not significant for this task, so the value from the equal variances assumed category was taken. Based on the results of the t-Test analysis, it does not indicate a statistically significant change ($p > 0.05$) for the 95% confidence interval. For this task, the null hypothesis cannot be rejected as the difference is not statistically significant.

Task 7. HR Student Employment

For this task, the focus changed from trying to access student resources to accessing faculty and staff resources, specifically the student employment page. With the framework enabled, took almost 2 clicks fewer to reach the desired page when the framework was enabled. Levene's test was significant for this task, the value from the unequal variances assumed category is taken. Looking at the Sig value for the t-Test analysis, it does indicate a statistically significant change ($p < 0.05$) for the 95% confidence interval. For this task, the null hypothesis can be rejected as the difference is statistically significant.

Task 8. HR Benefits

For this task, users were asked to navigate to the page that explained the benefits available to employees of the university. With the framework enabled, users were able to navigate to the page an about 0.5 clicks fewer using the framework. Levene's test was not significant ($p > 0.05$) for this task, so the value from the equal variances assumed category was used. Looking at the Sig value for the t-Test analysis, it does not indicate a statistically significant change ($p > 0.05$) for the 95% confidence interval. For this task, the null hypothesis cannot be rejected as the difference is statistically significant.

7.2.3.2 Navigation Summary

The focus of this research was the number of navigation steps taken to reach a desired page from the home page. Looking at evaluation of the responses to the survey, it appears that the framework had a slight effect on increasing overall usability. The null hypothesis was rejected with statistical significance on six of the eight tasks, and the two others appear to be better, despite the lack of statistical significance. Data from users who skipped tasks was unable to be used, but for this research there were only a few skips for each of the questions. The response numbers are close to 1% of the population of the University. While there was enough data to test for statistically significant differences between the two versions, it would have been better to test the framework on the production site for UNF. The real site garners far more traffic than the cloned site and would have given a truer indication of the success or failure of the framework. Despite

the lack of large amounts of data, the statistical analysis gives a good indication of success for the framework.

7.3 Survey Responses

In addition to collecting demographic and task completion data, users rated their agreement on various questions regarding their feeling towards the site. For the full list of questions, see Appendix B. The same questions were given to both groups of volunteers, regardless of the version of site they were presented with. These qualitative metrics, while not directly indicating the success or failure of the framework, help to better understand user attitudes to each site.

7.3.1 Survey Response Data

Each of the volunteers using the site was asked to gauge their feelings toward their version of the site on the eight questions listed in Appendix B. Some of these questions were taken from the Site Usability guide provided by the US government as standard for web design. Each of the eight questions was focused on the enjoyability and ease of use felt when navigating the respective site. The responses were all based on a 5-point scale, with 1 being strongly disagree and 5 being strongly agree.

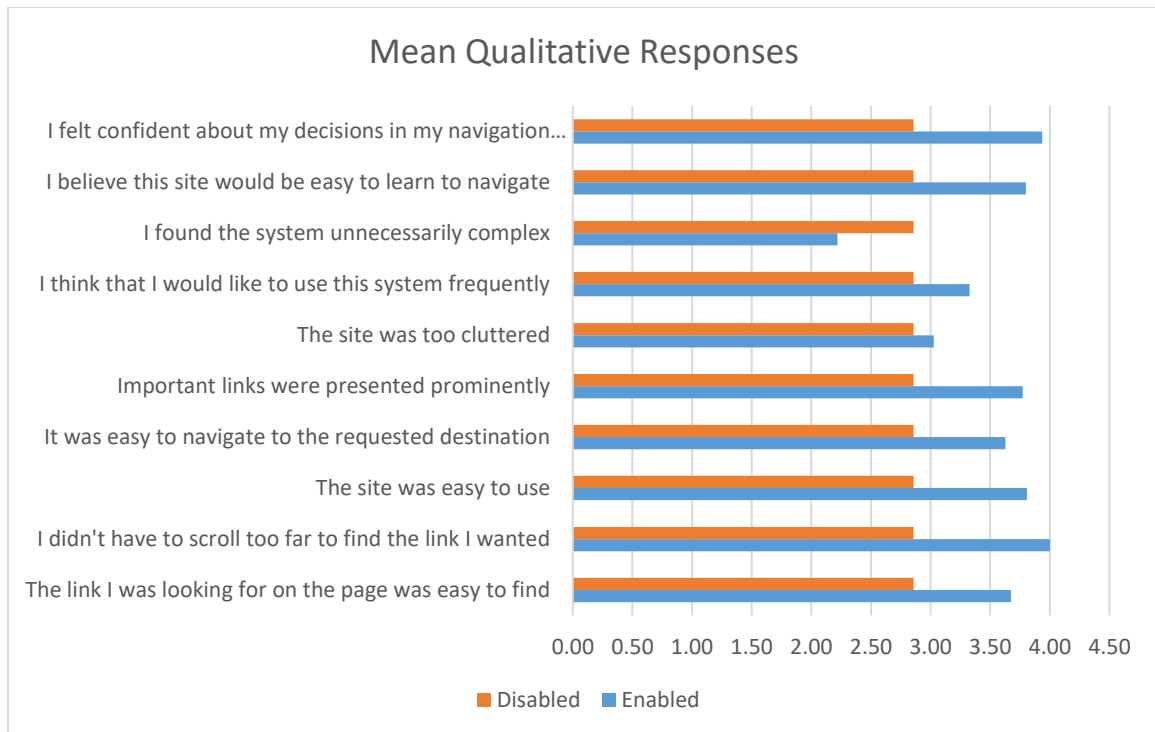


Figure 16 Qualitative Survey Responses

Based on the responses shown above, users of the framework enabled site had a more enjoyable time and felt like they could find what they were looking for more easily. On question 3 regarding how complex the system was, users felt the regular site was more complex. The final three questions all focused on the ease of navigation and were strongly in favor of the framework-enabled site. Each of these responses was not mandatory for completion, and some users elected not to answer all question. No statistical analysis was conducted as these were not the focus of this research, but rather, gave insight into the intangibles of the framework.

7.4 Prior Works on Active Analytics

This framework is an extension of research conducted by Will Carle for the University of North Florida (2016). That research focused on adapting webpages based on the current trends of most popular web pages based on Google analytics data, as a more real-time focused adaptation of user interfaces. While the research goal was not to compare these results to determine if the framework was an improvement over previous research, the framework does present an alternative way to dynamically adapt web pages to increase ease of use. To understand how the two versions of modification relate, a comparison the results to the previous research is provided below.

Of the eight tasks assigned to users to complete during testing, four were used by the previous research in a different order: admission deadlines, tuition and fees, graduate school programs, and employment benefits. In Carle's research, they conducted the same t-Test analysis to determine if their framework improved usability in terms of both time and number of steps. For the time-taken statistic they were only able to prove statistical improvement for one of their assigned tasks which directed users to a UNF library page, a task not included in this research (Carle, 2016). As shown previously, the framework was able to statistically improve two of the eight tasks in terms of time taken.

Similarly, in both sets of research, the mean number of steps it took to reach the desired page was improved. In the previous research, Carle was able to prove statistical

significance in two of the tasks, one of which was also included in this testing (Carle, 2016). The task four in this research and their final task directed users to the tuition and fees page for the University. Carle was able to reject the null hypothesis for this task while this data was unable to statistically improve this task with this data. The other task was again directing users to a Library page which was not included here. This research was able to increase the number of statically improved tasks from two to five while maintaining an overall improvement to the mean number of steps.

Chapter 8

FUTURE IMPROVEMENTS

This research has implications for real-world applications, both commercial and informational. While this research was limited in scope to the last year, a few major concepts were combined that could be broken out on their own as research. There was not time to flesh out all possible options and iterations of the framework but believe these ideas can be used to further the goal of increasing usability through anticipating user needs.

8.1 Features

This research was narrow in scope with respect to the amount of data taken from Google Analytics; only the page views for each of the sites was used to get the popularity ranking relative to a given timeframe. As stated previously, there are hundreds of data points gathered by Google Analytics and stored with each URL. One of the features considered, but ultimately decided against for the current scope, was including additional metrics such as length of time spent on the page and the exit rate. The exit rate could have been used as it is the percentage of users who leave a domain, unf.edu for example, and could be an indicator of finding what they were looking for. Future research could investigate combining various metrics and using weights to assign importance to each metrics and test the ability to increase overall user performance.

8.2 Timeframe Analysis

The core of this implementation was looking at the prior year's data and using that as a prediction model for the current timeframe. However, as noted in earlier chapters, there is a wealth of historical data spanning more than 10 years. Compared to that timeframe, the selection of only the last year represents less than 10% of the total historical data. Further research could be conducted in two ways. The first way would be to look at combining more than just the previous year's data, taking two, three, or more years, combining their number and assigning some weighted average to determine what links to promote on the site. There could be a large research push in this area, specifically, for example, research could be focused on just what that weighted average would be and analyzing the accuracy of it.

The second area of research that could be added on to this would take place at least one year after implementation. Since highlighting is applied to what it anticipated to be the most popular links, this could skew future results. Any modification to the homepage to entice users to click on a specific link will likely lead to increased traffic to that site.

Research could be conducted to understand and account for this skewing of results so that after the initial year, links are still being promoted by their relative importance, not just click length. This can be combined with the features talked about in section 8.1 to develop a more holistic view of site ranking as opposed to merely clicks.

8.3 Implementation

Finally, this research had User Interface element enhancements, in highlighting links to promote them, but it was not the focus of this research. Color and design theories have been explored in other areas to promote interaction but was limited to two colors and only highlighting hyperlinks in this research. The page used also contained elements like dropdown menus and scrolling news feeds that were not addressed in these modifications. Additional testing could be done to evaluate the effectiveness of promoting predicted links in a dropdown menu format the most popular sites.

Another issue with this implementation is that the site was a clone of the live site at a given time. During the snapshot window, the UNF site was in the middle of rebranding, changing the design and feel of pages within the site. While not intentional, the tasks, if followed in the least number of steps, did not include navigating to many, if any, of the newly redesigned sites. It would be worth investigating the ability of the framework to work once the redesign was finished depending on what elements they chose to update. While not complete during the snapshot, and modification or redesign of the homepage following the snapshot would have had a major implication on the testing as users would have a cognitive disconnect between what essentially would be an “archived” version of the site.

Overall, the implementation satisfied most of the core ideas of focus when research began. Access to production data from Google Analytics through a cooperation with

UNF was immeasurable helpful and real results were gathered. The framework has strong real-world implications and could be used by other universities and businesses to better promote information on their sites.

Chapter 9

CONCLUSION

The goal of this research was to develop a framework by which sites can automatically and adaptively change based on the usage trends during a given timeframe. To achieve this goal, a fully functioning framework was built that gathered data from historical sources and provided navigational element adaptations to users. Help was solicited from the UNF community in testing the proposed framework to determine its effectiveness. This system used real statistics from Google Analytics archives and presented users with popular sites for a given period. What was found is that the framework was able to statistically reduce the number of steps taken to reach a desired page in almost all tested scenarios. This is a good indication of the success of the framework and the overall usefulness it could provide to businesses.

The door has been opened to many more areas of research in terms of historical usage trends for predicting site navigation. The focus of this research was solely on page views as the metric of popularity, but there is a greater wealth of information available and numerous metrics stored for every site visit. Using these data points, research can be conducted into understanding how multiple metrics can, either individually or in combination, indicate user satisfaction and page popularity. With these ranking, developers could work to adaptively present the most likely pages to users in a variety of ways.

This framework has real-world implications and benefits, but mainly to sites that already have a wealth of analytics data. The reliance on historical data predicates the need for this data to exist before implementation. For those sites that have this data available, it has been shown that this framework can dramatically improve site usability.

REFERENCES

- Bhattacharya, Tanya, et al. "Web Usage Mining and Text Mining in the Environment of Web Personalization for Ontology Development of Recommender Systems." *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2016, pp. 78–85., doi:10.1109/icrito.2016.7784930.
- Büchner, Alex G., and Maurice D. Mulvena. "Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining." *ACM SIGMOD Record*, vol. 27, no. 4, Jan. 1998, pp. 54–61., doi:10.1145/306101.306124.
- Devanshu Dhyani, Wee Keong Ng, and Sourav S. Bhowmick. 2002. A survey of Web metrics. *ACM Comput. Surv.* 34, 4 (December 2002), 469-503.
- Doddegowda, B J, et al. "Extraction of Behavioral Patterns from Pre-Processed Web Usage Data for Web Personalization." *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2016, pp. 494–498., doi:10.1109/rteict.2016.7807870.
- Gaur, Loveleen, et al. "Google Analytics." *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies - ICTCS '16*, 2016, doi:10.1145/2905055.2905251.
- George, D., & Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference*, 17.0 update (10a ed.) Boston: Pearson.
- Geng, Ruili, and Jeff Tian. "Improving Web Navigation Usability by Comparing Actual and Anticipated Usage." *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 1, 2015, pp. 84–94., doi:10.1109/thms.2014.2363125.
- Gomathi, S., et al. "Web Page Design Optimisation Based on Segment Analytics." *2017 2nd International Conference on Computing and Communications Technologies (ICCCT)*, 2017, pp. 365–369., doi:10.1109/iccct2.2017.7972310.
- Goyal, Poonam, et al. "Designing Self-Adaptive Websites Using Online Hotlink Assignment Algorithm." *Proceedings of the 7th International Conference on Advances in Mobile Computing and Multimedia - MoMM '09*, 2009, doi:10.1145/1821748.1821860.

Hevner, A., S. March, J. Park, and S. Ram (2004) Design science in information systems research, *MIS Quarterly* 28 (1), pp. 75–105.

Issa, Tomayess, and Pedro Isaias. *Sustainable Design*. Springer London, 2015.

Khoo, Michael, et al. “Using Web Metrics to Analyze Digital Libraries.” *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries - JCDL '08*, 2008, pp. 375–384., doi:10.1145/1378889.1378956.

Kiseleva, Julia, et al. “Discovering Temporal Hidden Contexts in Web Sessions for User Trail Prediction.” *Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion*, 2013, doi:10.1145/2487788.2488120.

Ma, Jiangang, et al. “Efficiently Finding Web Services Using a Clustering Semantic Approach.” *Context Enabled Source and Service Selection, Integration and Adaptation*, 22 Apr. 2008

Mcgrath, Owen G. “Insights and Surprises from Usage Patterns.” *Proceedings of the 36th Annual ACM SIGUCCS Conference on User Services Conference - SIGUCCS '08*, 2008, doi:10.1145/1449956.1449976.

Norguet, Jean-Pierre, et al. “Semantic Analysis of Web Site Audience.” *Proceedings of the 2006 ACM Symposium on Applied Computing - SAC '06*, Apr. 2006, pp. 525–529., doi:10.1145/1141277.1141401.

Peppers, Ken, et al. “A Design Science Research Methodology for Information Systems Research.” *Journal of Management Information Systems*, vol. 24, no. 3, Jan. 2007, pp. 45–77., doi:10.2753/mis0742-1222240302.

Piao, Bin, and Yuzuru Tanaka. “Interactive Framework for Exploratory Search, Integration, and Visual Analysis of Semantic Web Resources.” *Proceedings of the 16th International Conference on Information Integration and Web-Based Applications & Services - IiWAS '14*, 2014, doi:10.1145/2684200.2684351.

Reddy, K. Sudheer, et al. “Understanding the Scope of Web Usage Mining & Applications of Web Data Usage Patterns.” *2012 International Conference on Computing, Communication and Applications*, 2012, doi:10.1109/iccca.2012.6179230.

Rodden, Kerry, et al. “Measuring the User Experience on a Large Scale.” *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*, Apr. 2010, pp. 2395–2398., doi:10.1145/1753326.1753687.

Sajeev, G P, and P T Ramya. “Effective Web Personalization System Based on Time and Semantic Relatedness.” *2016 International Conference on Advances in Computing*,

Communications and Informatics (ICACCI), 2016, pp. 1390–1396.
doi:10.1109/icacci.2016.7732242.

Salini, Andrea, et al. “Leveraging Web Analytics for Automatically Generating Mobile Navigation Models.” *2016 IEEE International Conference on Mobile Services (MS)*, 2016, pp. 103–110., doi:10.1109/mobserv.2016.25.

Salkind, Neil. “t Test, Independent Samples.” *Encyclopedia of Research Design*, 2010, pp. 1551, SAGE Publications

Sathiaseelan, J. G. R., et al. “Multi-Level Secure Framework (MLSF) for Composite Web Services.” *Proceedings of the 2nd International Conference on Interaction Sciences Information Technology, Culture and Human - ICIS '09*, 2009, doi:10.1145/1655925.1656030.

Sleeper, Manya, et al. “Exploring the Benefits and Uses of Web Analytics Tools for Non-Transactional Websites.” *Proceedings of the 2014 Conference on Designing Interactive Systems - DIS '14*, 2014, doi:10.1145/2598510.2598555.

Zheng, Guangzhi, and Svetlana Peltsverger. "Web Analytics Overview." *Encyclopedia of Information Science and Technology*, Third Edition (2015): 7674-683. Web.

APPENDIX A

NAVIGATION TASKS

1. Task #1 Colleges And Programs: A prospective student, Bob, is interested in attending UNF beginning in the summer semester. He is undecided in what he wants to study, so help Bob see a list of Colleges and Programs available at UNF. Please navigate to the "Colleges and Programs" page.
 - a. Options: [SKIP]
 - b. Completion Message: Completion message: Great! You've completed the task, click [Next] to start the next task

2. Task #2 Course Catalog: Bob, after deciding on two possible schools he is interested in, would like to know more about the classes offered to students. Please navigate to the "Course Catalog for Undergraduate Students" page
 - a. Options: [SKIP]
 - b. Completion Message: Completion message: Great! You've completed the task, click [Next] to start the next task

3. Task #3 Graduate School Programs: Bob has also considered pursuing a Masters degree following his undergraduate degree. He would like to know more about the programs offered to see if any exist in the two schools he is considering. Please navigate to the "Graduate Programs at UNF" page.
 - a. Options: [SKIP]
 - b. Completion Message: Completion message: Great! You've completed the task, click [Next] to start the next task

4. Task #4 Tuition: Bob will likely be paying his way through school and would like to obtain information on tuition and fees for UNF students. Help Bob by navigating to controller page that displays tuition and fees details. Please navigate to the "Tuition" page with the breakdown of tuition and fees for students.
 - a. Options: [SKIP]
 - b. Completion Message: Completion message: Great! You've completed the task, click [Next] to start the next task

5. Task #5 Admissions Deadlines: Bob wants to begin school in the Summer and would like to know information regarding application deadlines. Help Bob by

navigating to UNF admissions page that displays deadlines information. Hopefully he won't need help applying too. Please navigate to the UNF Admissions "Deadlines" page

- a. Options: [SKIP]
- b. Completion Message: Completion message: Great! You've completed the task, click [Next] to start the next task

6. Task #6 Apply: Bob believes he has decided on a major to pursue and wants to know more about the Application process and possibly apply today. Help Bob by navigation to the UNF admissions page that displays the steps for applying.

Please navigate to the "Apply Now" page

- a. Options: [SKIP]
- b. Completion Message: Completion message: Great! You've completed the task, click [Next] to start the next task

7. Task #7 HR Employment: Bob, an enterprising young man, wants to work his way through school. He is interested in learning about student employment opportunities. Please help Bob find the Human Resources page that lists the types of student employment at UNF. Please navigate to the "UNF Student Employment" page

- a. Options: [SKIP]
- b. Completion Message: Completion message: Great! You've completed the task, click [Next] to start the next task

8. Task #8 HR Benefits: Bob has some questions on benefits offered to UNF employees. Help Bob by navigating to human resource page that displays benefits information. Please navigate to the Human Resources "Benefits" page

- a. Options: [SKIP]
- b. Completion Message: Completion message: Great! You've completed the task, click [Next] to start the next task

APPENDIX B

USER EXPERIENCE SURVEY

Demographic Questions:

Age

- 18-25
- 26-35
- 36-45
- 46-55
- 56-65
- 65+

What is your current class standing?

- Non-matriculated
- Freshman
- Sophomore
- Junior
- Senior
- Graduate

How experienced are you in using the internet?

- Very Experienced
- Some Experience
- Limited Experience
- No Experience

Which browser did you use to view the site?

- Internet Explorer
- Google Chrome
- Safari
- Firefox

Is English your primary language?

- Yes
- No

Task Specific Questions:

Have you visited the UNF website before?

- Yes

- No

If yes how often do you visit the UNF website?

- A few times a year
- A few times per month
- Once a week
- Multiple times a week
- Daily
- Multiple times a day

Were you able to complete all the tasks?

- Yes
- No

If not, why were you not able to complete the tasks?

Did you get lost at any point while trying to complete a task?

- Yes
- No

If yes please describe what happened

Were you frustrated at any point when trying to complete a task?

- Yes
- No

If yes please describe what caused the frustration.

User Experience Ratings:

Please rate the following statements from 1 (strongly disagree) to 5 (strongly agree):

	Strongly Agree (5)	Agree (4)	Neither Agree nor Disagree (3)	Disagree (2)	Strongly Disagree (1)
The link I was looking for on the page was easy to find					
I didn't have to scroll too far to find the link I wanted					
The site was easy to use					
It was easy to navigate to the requested destination					
Important links were presented prominently					
The site was too cluttered					

	Strongly Agree (5)	Agree (4)	Neither Agree nor Disagree (3)	Disagree (2)	Strongly Disagree (1)
I think that I would like to use this system frequently					
I found the system unnecessarily complex					
I believe this site would be easy to learn to navigate					
I felt confident about my decisions in my navigation choices					

APPENDIX C
IRB DOCUMENTS

IRB Approval Letter



MEMORANDUM


DATE: March 8, 2018

TO: Mr. Jacob Koza

VIA: Dr. Karthikeyan Umopathy
Computing

FROM: Dr. Jennifer Wesely, Chairperson
On behalf of the UNF Institutional Review Board

RE: Review of New Project by the UNF Institutional Review Board
IRB#1183620-1: "Predictive Analytics: Suggesting Navigational Links to Users based on
Temporal Analytics Data"

UNF IRB Number: 1183620-1
Approval Date: 03-08-2018
Expiration Date: 03-08-2019
Processed on behalf of UNF's IRB 

This is to advise you that your project, "Predictive Analytics: Suggesting Navigational Links to Users based on Temporal Analytics Data" underwent "[Expedited](#)" [Category 7](#) review on behalf of the UNF Institutional Review Board. Your reviewer recommended approval without further modifications.

A waiver of *signed* informed consent was requested and approved for this research based on the criteria in [45 CFR 46.117\(c\)](#). You may use the electronic consent procedures for participants as outlined in your approved documents. Only the approved versions of the consent information should be used during this research. All participants must receive a stamped and dated copy of the approved informed consent document when possible.

This approval applies to your project in the form and content as submitted to the IRB for review. Any variations or modifications to the approved procedures or documents must be cleared with the IRB prior to implementing such changes. *For example*, if you plan to make changes to your stamped and dated informed consent form, it will be necessary to submit a copy of the revised form via an amendment so that it can be reviewed and approved prior to use. Once approved, a new stamp and date will be included on the revised consent form so that it can be used. To submit an amendment, please complete an [Amendment Request Document](#) and submit it along with any updated documents affected by the changes via a new package in IRBNet. Any unanticipated problems involving risk and any occurrence of serious harm to subjects and others shall be reported by completing this [Event Report Form](#) and sending it promptly to the IRB within 3 business days.

Your study has been approved for a period of 12 months as of 03/08/2018. If you would like your project to continue for more than one year, you will be required to provide a completed [Status Report](#) and other continuing review documentation to the UNF IRB prior to **02/08/2019**. An extension will be necessary if your study will be continuing past the 1-year anniversary of the approval date. *We ask that you submit your status*

report and other continuing review information 30 days before the expiration date as noted above to allow time for review and processing. When you are ready to close your project, please complete a [Closing Report Form](#). Please note that it will be necessary to create a new package in IRBNet in order to submit amendments, status reports, or closing reports in the future. All applicable records relating to this research shall be retained for at least 3 years after completion of the research.

CITI Training for this Project:

Name	CITI Expiration Date
Mr. Koza	01/15/2021
Dr. Umapathy	07/12/2018

CITI Course Completion Reports are valid for 3 years. Please note that Dr. Umapathy's CITI completion report will expire this year. Please renew CITI training before your current training expires. Although CITI sends out reminders 90 days prior to expiration, it is the investigator's responsibility to complete the refresher course when it becomes available. You can access the CITI refresher course by following this link: <http://www.citiprogram.org/>. Should you have questions regarding your project or any other IRB issues, please contact the research integrity unit of the Office of Research and Sponsored Programs by emailing IRB@unf.edu or calling (904) 620-2455.

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within UNF's records. All records shall be accessible for inspection and copying by authorized representatives of the department or agency at reasonable times and in a reasonable manner. A copy of this approval may also be sent to the dean and/or chair of your department.

VITA

Jacob Koza has a Bachelor of Arts degree from Mercyhurst University in Intelligence Studies and expects to receive a Master of Science Degree in Software Engineering from the University of North Florida. Jacob is currently employed as a Software Developer at TIAA Bank and has worked previously as an Application Systems Specialist and Analyst at the University of North Florida. Jacob has been working full-time as a software developer for close to two years.

Jacob has interests in developing custom web and mobile applications, focused on the Microsoft and Android development platforms. He has experience working on the full Microsoft stack, from ASP to MVC, as well as emerging technologies such as ReactJS. He has exposure to a wide variety of technologies and languages and strives to continually learn about new and emerging technologies. For more information about him, you can visit his website at <http://www.jacobkoza.com> to view his full biography and ever-expanding portfolio of work.