EPiC Series in Computing

Volume 69, 2020, Pages 32–39

Proceedings of 35th International Conference on Computers and Their Applications

NRDC Data Visualization Web Suite:
Tool for Data Visualization, Comparison, and Prediction Analysis

Andrew Muñoz, Frederick C. Harris, Jr., and Sergiu Dascalu
University of Nevada, Reno
amunoz24@nevada.unr.edu, fred.harris@cse.unr.edu,
dascalus@cse.unr.edu

Abstract

The Nevada Research Data Center (NRDC) is a research data management center that collects sensor-based data from various locations throughout the state of Nevada. The measurements collected are specifically environmental data, which are used in cross-disciplinary research across different facilities. Since data is being collected at a high rate, it is necessary to be able to visualize the data quickly and efficiently. This paper discusses in detail a web application that can be used by researchers to make visualizations that can help in data comparisons. While there exist other web applications that allows researchers to visualize the data, this project expands on that idea by allowing researchers the ability to not only visualize the data but also make comparisons and predictions.

1 Introduction

Big Data is a term that has become a staple in the field of Computer Science. It represents exceptionally large sets of data that need to be processed and analyzed to potentially reveal useful information such as patterns or trends. In [1] the authors stated that the mainstream definition of Big Data consists of the three V's, which are Volume, Velocity, and Variety. Volume refers to the amount of data being collected, Velocity refers to the speed of the data, and Variety refers to the various types of data which can consist of structured and unstructured data.

One important aspect in the processing of Big Data is the use of data visualization methods. Data visualization is the use of graphical formats such as charts and graphs to display data in a way that allows for readability and better understanding of the data [2]. In a previous project [3] conducted at the University of Nevada, Reno (UNR), an application was developed by a group of researchers to enable users to access and visualize data from the Nevada Research Data Center (NRDC) through the CHORDS visualization service. The use of CHORDS, which is a cloud hosted service, allowed users to visualize NRDC data sets from a web browser.

The NRDC is a cyber infrastructure data management center that is responsible for the collection, storage, querying and distribution of sensor data [4]. These sensors are located throughout the state of Nevada and collect environmental data that is used for cross-disciplinary research between many

different research centers. To this date, 15 environmental research sites in Southern Nevada have been connected to NRDC, each sending every five minutes data measurements collected from over 800 sensors. Currently, NRDC stores over 5.1 billion of environmental data records, as well as millions of images collected from related web cameras.

The work presented in this paper aims to create a data visualization tool that can be used by researchers to perform quick data visualization and analysis of NRDC data through an effective and accessible application. Through the application, users will be able to make simple visualizations, comparisons, as well as predictions on datasets that will permit detailed analysis of the data.

This paper is structured as follows. In Section 2, similar work is referenced and discussed. Section 3 gives an overview of the problem this project addresses and explains the software design. Section 4 goes into detail about the implementation of the project. Section 5 discusses the overall performance results and evaluation of the final project prototype. Lastly, Section 6 presents concluding thoughts on the project prototype as well as several potential future developments

2 Related Works

This section provides a brief overview on various related works. These works are broken down between two subsections with the first detailing related applications and the second explaining some related studies that focus on the combination of data science and software engineering.

2.1 Similar Applications

As mentioned previously, a related application was developed by a group of student researchers at UNR which allowed users to visualize NRDC data through a cloud hosted web portal [3]. The system developed provided users with the ability to select measurement tools on the network and stream real-time data into a graph. The authors state that the project was developed to act as a middle layer between the NRDC database and the CHORDS service that could then be used to view data in a visualized format.

Another similar application was also developed at UNR and is similarly closely connected to the NRDC. In this paper [5], the project presented is a mobile application that enables the collection of metadata from the NRDC and provides useful tools that can be used by technicians for quality assurance of the sensor networks. Similar to the previous application mentioned, this application also allows for near-real time collection of the metadata. This project aids in data management capabilities as it provides researchers and technicians with an accessible application that syncs directly to the database to make datasets easy to find and maintenance easy to perform.

2.2 Related Studies

There have been many different studies that tie closely to the topics of big data and software engineering. One such study conducted by Martinez-Fernandez et. al[6] attempts to understand whether the use of quality models in software analytics tools provides reliable and pertinent information about the quality of a process and whether or not it is useful for professionals in the field. The study conducted assessed four companies over the span of a year to determine if such a tool was useful for the companies. It was determined through both quantitative and qualitative results that the tool elicited mostly positive results and was found to be very useful and reliable.

In another article [7] the authors detail some of the primary challenges faced by software engineers when developing big data systems. The article consists of summaries that span across 26 different challenges that the authors identify in the development of big data systems. Following the identification of the 26 challenges, the authors were able to categorize the challenges based on the

phases of the software development lifecycle. It is stated that the main purpose of the article was to identify challenges to make them known and possibly allow for some insight into potential solutions to these challenges.

An article by Lin and Huang [8] presents a new software engineering lifecycle process designed by the authors specifically for big data projects. It is noted in the article that there is no set standard for big data projects and since it is an ever-growing topic, a new lifecycle process is essential to aid in development. The process described in the paper takes into account four main elements which include data variety, data innovation, software engineering, and data analysis. The authors tested the process on big data projects in the banking and retail fields, but concluded that further evaluation was needed to determine the effectiveness of the new process.

3 Project Overview and Design

In this section, we describe how the NRDC Data Visualization Web Suite has been developed, including the problem addressed and software development tools used. In the design process, requirements and diagrams were made to help provide an overall structure to the use cases and the design of the application being developed. In order to fully examine the development process, this section is broken down into three separate subsections, namely the Problem, Software Design and Technology, and Detailed Design.

3.1 The Problem

There have previously been projects that were developed to aid in the visualization and access of data. The main problem addressed in our project is that of increased accessibility and readability. The application itself is designed to allow users to search for specific datasets and create graphical data visualizations that will aid in data analysis. Currently, existing solutions are missing the ability to allow users to compare different datasets or make prediction analysis. Our proposed solution addresses this problem and attempts to provide the answers to allow users the opportunity for a more detailed analysis.

3.2 Software Design and Technology

The project was developed as a web application to make it easily accessible for users. The primary software tools that were used during development are listed as follows:

1. Visual Studio [9] - Developed by Microsoft, Visual Studio is a integrated development environment (IDE) that can be used to develop various programs such as mobile applications, virtual reality applications, and web applications.
2. .NET Framework [10] - Software developed by Microsoft, .NET is a framework that is primarily used to build web applications and services that can run cross-platform but works best with windows. It works well with Visual Studio as a package that can be installed.
3. C# Language [11] - Programming language that is similar in some ways to C and C++ languages. It is an object-oriented language that is useful for developing robust applications.
4. HTML [12] - Stands for Hyper-text Markup Language. It is the standard language used for designing pages that are displayed on a web application. It can be used in tandem with other technologies and languages such as CSS and JavaScript.

3.3 Detailed Design

The activity diagram shown in Figure 1 displays the main flow of the NRDC Data Visualization Web Suite. Upon starting the application, the user is presented with the home page, which allows the user to select a method of analysis. These methods will include visualization, comparison, and prediction. Once a method has been selected, the user will input information that will allow him or her to query a specific dataset from the database. The user is then required to input specific data such as site to be selected, environmental parameters of interest, and date range among others. Next, the user can press a button that allows him or her to see the results of their query in a visualized format. If the users wish to refine their query, they may click a button below the graph to make changes. Finally, if a user finds the visualization useful he or she may save it to his or her profile for future reference and retrieval.

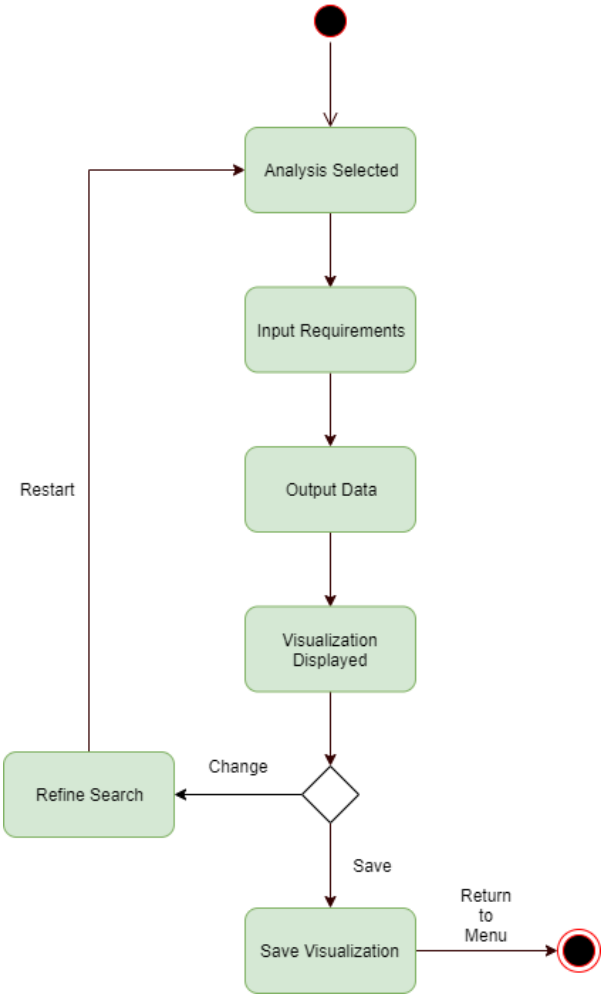

Figure 1: Activity diagram that displays the main flow of the web application.

4 Application Prototype

The application was implemented using the software tools mentioned in Section III-B. Through these tools, the implementation of the project was very simple. The project itself is very simple in concept, as it is a data visualization tool that allows users to make detailed and thorough analysis on a specific dataset. The prototype of the application is designed to be very simple and easy to use. This allows for learn ability and efficiency as users are not confused when using the application. When a user begins the application, they will be greeted by a login screen as displayed in Figure 2. Here the user will be asked to input their credentials or be able to create an account. Once the user is logged in, they will then be will be taken to the home page where they will have a variety of options from which they can choose. The user will be able to manage account settings such as view history, sign out, and be able to use the primary visualization services.

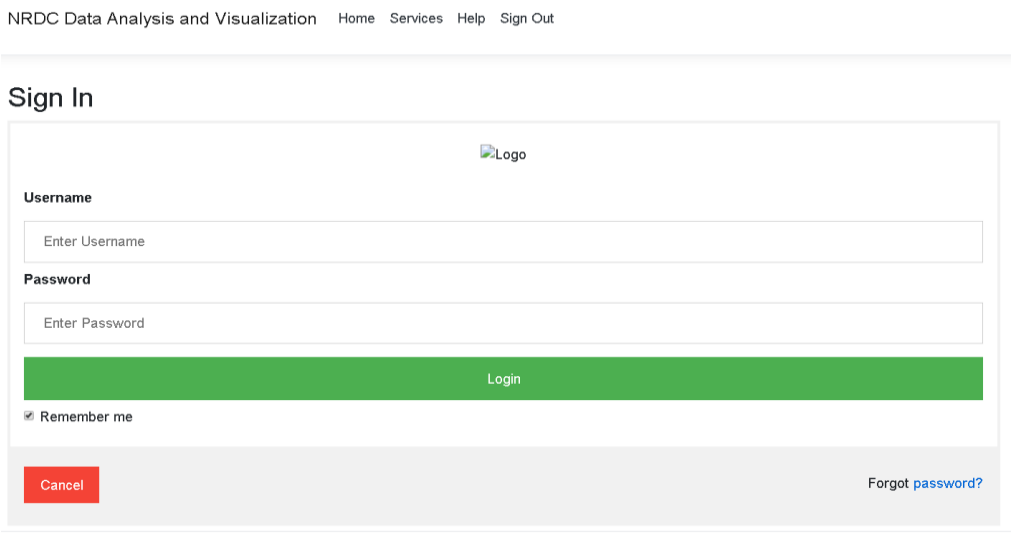

Figure 2: Screenshot displaying the login page the user sees upon entering the application.

As previously stated, the services that the application provides include visualization, comparison, and prediction. The user is able to choose one of these services from the home screen. Once a service is selected, the user is taken to a form-like screen where they can input the information needed to query the data. Figure 3 shows the page a user will see when attempting to create a comparison between datasets. The inputs are provided through specific drop-down menus that allow users to select the available variables.

At this time, there are three parameters users can choose from, specifically temperature, precipitation, and solar radiation. Currently, the users have the possibility to access four research sites from which they can choose to obtain data. However, since NRDC contains data from 15 sites and the measurements are from about 70 different types of sensors, both additional sites and other parameters will be made available in the future through the NRDC Data Visualization Web Suite that we have developed.

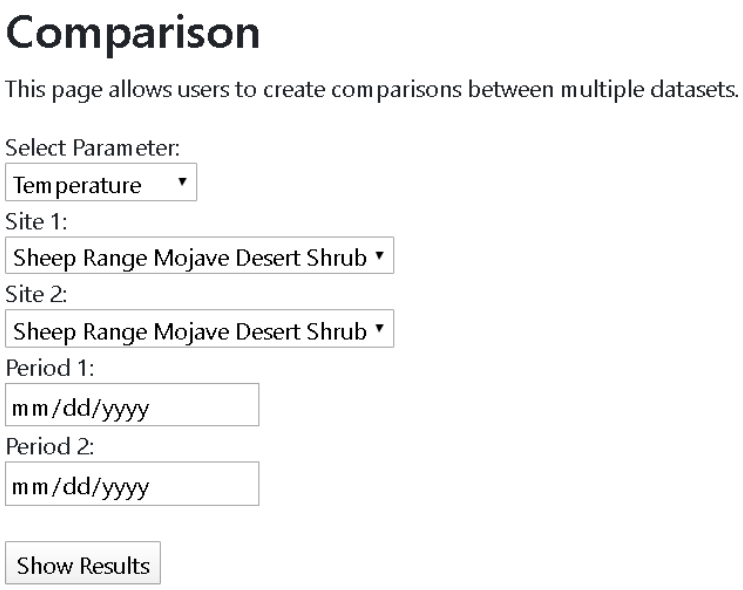

Figure 3: User interface for users to input information for a query to create a comparison between datasets.

In terms of dates, the users can compare between two different periods from which they can obtain two sets of data visualizations. Additionally, each data input page contains different information that is required for a successful query. The visualization page for instance requires the users to input a visualization type, i.e. line graph or bar graph. The prediction page requires users to input the type of interval, backward or forward interval, as well as select a prediction algorithm. Once the input values have been selected, the user can then create a visualization graph. Figure 4 shows the results page from a query for visualization. The results shown here are in the form of a line graph for temperature. The users also have the option to refine their search or save the visualization and query to their profiles.

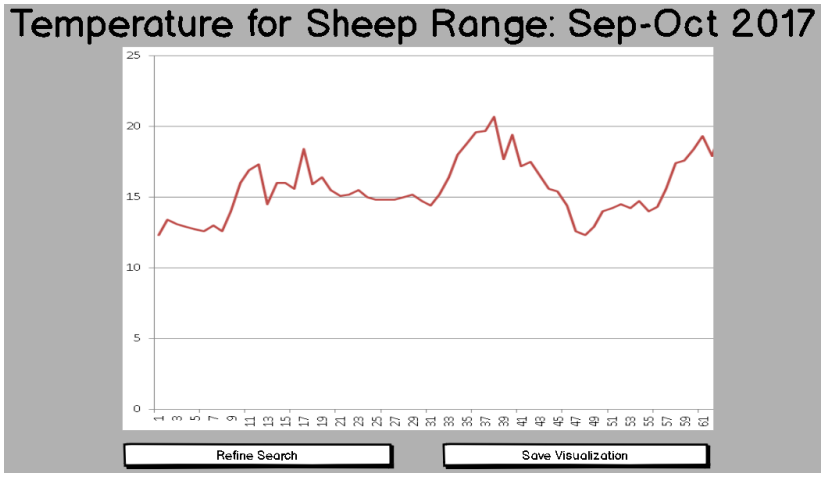

Figure 4: Screenshot of the results page with a visualization graph spanning two months of temperature data.

5 Results and Evaluation

Following the implementation of the project, the prototype needed to be prepped for demo and initial user testing. The main focus made in terms of the demo were on usability and learning ability. It was important to develop an application that provided users with an easy and effective means to visualize data from the NRDC. The results of the demo were mostly that the application provided a nice flow and organic layout that made it easy to use. The variety of services also provided users with a fairly good amount of options that allowed them to perform detailed analysis on any given dataset. Of interest to users have been all three major pieces of functionality, namely visualization, comparison, and prediction.

At this time the project is not as fleshed out as originally intended, but still provides functionality and ease of use. One of the main issues with the application was that it was a bit basic and still needed some visual improvements. The application as it stands is functional but lacks in terms of user interface alternatives. The second issue was with the user profile as it was also simplistic and lacking some features that would make it more robust. Such features could include an actual profile page and preferences. While the project is not as robust as intended, it provides a simple and functional prototype that shows the main features of application.

6 Future Work and Conclusions

This paper has described the development and implementation process of the NRDC Data Visualization Web Suite. The main purpose of the project is to develop a software interface that provides users with the ability to easily and effectively create data visualization graphs and charts. In addition, users are also able to perform comparison and prediction analysis on selected datasets. This added functionality is currently lacking in existing solutions and makes this application relevant.

As for future work, the application needs some refinements as well as additional features that will make it more aesthetically and functionally appealing. First, the application should be made to provide the user with more options in terms of preferences and settings. Therefore, the application should be able to switch between a dark and light theme as currently only a light theme is available. Second, the application needs to be fully linked to the NRDC database to obtain full datasets. Due to a short development period, it was not possible to make the connection to the database so several datasets in the form of CSV files were used to demonstrate the designed functionality. Third, the application could also be developed as a mobile application to increase accessibility.

In conclusion, the NRDC Data Visualization Web Suite developed is functional and accomplishes its main core services. However, the application is a bit limited in terms of user options and needs more functionality in order to increase its usefulness and effectiveness. These additions are planned to be implemented, tested and integrated in January and February 2020. We also plan to conduct a full user study to determine, in principal, the application’s effectiveness and ease of use. This will be done once the user interface will be further enhanced and some more functionality will be added.

References

[1] SAS Institute, Inc., "What it is and Why it Matters?," First accessed December 15, 2019. [Online]. Avaliable: https://www.sas.com/en_us/insights/big-data/what-is-big-data.html.

[2] Tableu Software, "Data Visualization Beginner's Guide: A Definition, Examples, and Learning Resources," First accessed December 15, 2019. [Online]. Available: https://www.tableau.com/learn/articles/data-visualization.

[3] P. Jirasessakul, Z. Waller, P. Marquis, V. Le, C. Scully-Allison, S. Strachan, F. C. Harris, Jr., S.M. Dascalu, "Generalized Software Interface for CHORDS," in *ISCA 27th International Conference on Software Engineering and Data Engineering (SEDE 2018)*, 2018.

[4] University of Nevada, Reno., "Nevada Research Data Center: Streaming Data Management for Sensor Networks," First accessed October 6, 2019. [Online]. Avaliable: http://www.sensor.nevada.edu/NRDC/.

[5] C. Scully-Allison, H. Munoz, V. Le, S. Strachan, E. Fritzinger, F. C. Harris, Jr., S.M. Dascalu, "Advancing Quality Assurance Through Metadata Management: Design and Development of a Mobile Application for the NRDC," in *IJCA International Journal of Computers and Their Applications*, Vol.25, No. 1, pp. 20-29, 2018.

[6] S. Martinez-Fernandez, A. M. Vollmer, A. Jedlitschka, X. Franch, L. Lopez, P. Ram, P. Rodriguez, S. Aarama, S. Bagnato, M. Choras, and J. Partanen, "Continuously Assessing and Improving Software Quality With Software Analytics Tools: A Case Study," *IEEE Access*, Vol. 7, pp. 68,219 - 68,239, 2019.

[7] O. Hummel, H. Eichelberger, A. Giloj, D. Werle, and K. Schmid, "A Collection of Software Engineering Challenges for Big Data System Development," in 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 362 - 369. 2018.

[8] Y. Lin and S. Huang, "The Design of a Software Engineering Lifecycle Process for Big Data Projects," *IT Professional*, Vol. 20, No. 1, pp.45-52, January 2018.

[9] Microsoft, "Visual Studio 2019," First accessed November 10, 2019. [Online]. Avaliable: https://visualstudio.microsoft.com/vs/.

[10] Microsoft: .NET Core 3.1.0, "What is .NET Framework," First accessed November 11, 2019. [Online]. Avaliable: https://dotnet.microsoft.com/learn/dotnet/what-is-dotnet-framework.

[11] Microsoft Docs, "A Tour of the C# Language," First accessed November 13, 2019. [Online]. Avaliable: https://docs.microsoft.com/en-us/dotnet/csharp/tour-of-csharp/.

[12] W3Schools, "Introduction to HTML," First accessed November 12, 2019. [Online]. Avaliable: https://www.w3schools.com/html/html_intro.asp.