A CONCEPTUAL MODEL AND WEB PLATFORM FOR DISCOVERING DIGITAL HEROES FROM STACK OVERFLOW DEVELOPER COMMUNITY FORUM USING A DATA MINING APPROACH

By

ABDUR ROB

A THESIS SUBMITTED TO THE SCHOOL OF INFORMATION SCIENCES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN INFORMATION TECHNOLOGY, DEPARTMENT OF INFORMATION TECHNOLOGY, MOI UNIVERSITY.

2018

DECLARATION

DECLARATION BY THE CANDIDATE:

This study is my original work and has not been presented for a degree in any other University. No part of this thesis may be reproduced without the prior written permission of the author and/or Moi University.

_____ _____

ABDUR ROB Date

IS/MSC/IT/02/16

DECLARATION BY SUPERVISORS:

This thesis has been submitted for examination with our approval as University Supervisors:

_____ _____

DR. JOHN K. TARUS Date

Department of Information Technology

Moi University, Eldoret, Kenya

_____ _____

DR. IRENE MOSETI Date

Department of Information Technology

Moi University, Eldoret, Kenya

DEDICATION

I would like to dedicate this thesis to my parents Abdur Razzak and Fatema Begum and my brothers and sister for their moral support and encouragement all through. May ALLAH bless and reward you all.

ACKNOWLEDGEMENT

I would like to express my deep gratitude to almighty ALLAH for giving me ability to work on this thesis. I wish to acknowledge the guidance, advice and supervision accorded to me as I wrote this thesis by my two supervisors Dr. John K. Tarus who works in the Directorate of ICT and Dr. Irene Moseti who is the head of the Information Technology Department, the criticism and comments given ultimately led to a successful thesis writing and defense.

Many thanks to Mr. Kiget, who is a lecturer in the department of Information Technology, for his suggestions and guidance related to the study. I would also like to thank Mr. Khaliqur Rahman, one of my senior brothers, for his moral support and motivation to writing this thesis.

I also thank to my scholarship host, the Commonwealth Scholarship and Fellowship Plan (CSFP) under the Association of Commonwealth Universities (ACU) for giving me this opportunity to travel all the way from my home country Bangladesh to Kenya and pursue my Master's degree from Moi University and learn about African culture. Without ACU it would have been impossible to me.

Lastly I would wish to appreciate the support of my friends for the discussions and criticism they gave as first audience of the thesis defense, thank you and I wish you all success.

ABSTRACT

Heroes are crucial in our general life. They inspire us to do good deeds. Likewise, in the digital world people are helping others using online community forums even without knowing the people they are helping. They are referred to as digital heroes and their efforts enrich the community forums and digital world day by day. The available literature does not reveal an existing platform for analyzing people's digital activities and identifying digital heroes in the digital world. This study aimed to establish digital hero selection criteria based on community forum activities using data mining approach and develop a web platform. The platform will aid in discovering and documenting digital heroes so that future generations can know about them and be inspired by their contributions to the digital community as well as be motivated to do good deeds. Using homogeneous purposive sampling technique, this study used a public dataset of 1,889,860 users between 2008 and 2017 from the Stack Overflow online developer community forum and analyzed using data mining approach in order to discover the digital heroes based on their digital activities and proposed selection criteria. The data mined was mainly user's information including user name, registration date, about, user id, total votes (up and down) and reputation number. The collected information was analyzed based on four proposed hero selection criteria for identifying digital heroes: Experience, Accuracy, Activity and Trust, which were established by collecting experts' opinion using questionnaires. In order to establish digital hero selection criteria, this study used convenience sampling technique to target 45 experts who know about the Stack Overflow community forum, understand how it works, and know about developer community members and their activities in the forum. Furthermore, this study developed a web platform referred to as Digital Hero Discovery (DHD) web platform that was guided by conceptual data analysis model to analyze, discover and document digital heroes. The study was able to discover a total number of 3,231 digital heroes who qualified in all the criteria using the Knowledge Discovery in Databases (KDD) process of data mining approach to analyze data as well as hero selection criteria. The digital heroes were classified into four proposed hero categories: A, B, C and D based on their contributions in the community. From the findings, the categories had the following number of heroes: A – 1; B – 45; C – 2,411; and D – 774. The proposed digital hero selection criteria and data mining approach would be the start of the journey of digital hero discovery to inspire digital world people to help others accurately and frequently. The developed DHD web platform could be used in other digital communities including academic, medical and social communities in order to discover digital heroes based on their contributions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

API	-	Application Program Interface
CSS	-	Cascading Style Sheets
DHD	-	Digital Heroes Discovery
HTML	-	Hypertext Markup Language
KDD	-	Knowledge Discovery in Databases
PHP	-	Hypertext Preprocessor
SO	-	Stack Overflow
SQL	-	Structured Query Language
MySQL	-	MySQL is a full-featured open source relational database management system (RDBMS) based on Structured Query Language (SQL)
KDD	-	Knowledge Discovery in Databases
SDLC	-	Systems Development Life Cycle
SPM	-	System Prototype Method
SSADM	-	Structured Systems Analysis and Design Methodology
LBMS	-	Learmonth Burchett Management Systems
CCTA	-	Central Computer Telecommunications Agency
DFD	-	Data Flow Diagram
ERD	-	Entity Relationship Diagram

CHAPTER ONE - INTRODUCTION

1.0 Introduction

Heroes are people who are willing to do anything to save other people and are never afraid to stand for something good (Miller, 2011). They have great capacity to face challenges and are willing to sacrifice for the sake of humanity. According to Dictionary.com (n.d.), a person who, in the opinion of others, has special achievements, abilities, or personal qualities and is regarded as a role model is ideally a hero. Reader's Digest Magazine (n.d.) acknowledges that a hero works beyond the scope of his or her job, possibly as a volunteer. A hero responds to a social need as well as the needs of a person or group. A hero is selfless, genuine and a good person who gets the undivided attention of all of us and causes change. Heroes are people who, through their actions, make the world better.

The digital world refers to an environment that interconnects people through the Internet, online media, and digital devices. The Internet has made it possible for people to connect whenever it is necessary. Digital life gives people opportunities to connect worldwide within a single moment. We can share our every activity to thousands of people worldwide. People are now engaged with digital world life just like in physical life. In summary, we can say that digital life is our online footprint.

In general life, heroes are those who have dedicated their life for the sake of humanity. Similarly, in digital life, a lot of people work to help millions of people worldwide without even knowing them. As we have communities in general life, likewise in digital life, there are individuals who respond to questions on online forums and communities to help people they do not know. Could these people who have worked for information revolution be part of digital heroes? Therefore, this study considers digital heroes as

people who dedicate their effort to help digital communities, by making an incredible positive impact to the digital world.

Though we have digital heroes in the digital world, there is no platform to analyze their activities and identify them and keep their records in one place for the future generation. The targeted population of this study is all the Stack Overflow[1] community forum users. The data considered is all the questions asked and replies given on the forum. For the sake of data analysis and developed data processing model, this study used the developer community domain only. However, the developed data processing platform can be used for any other domain by changing variables based on that domain. Therefore, this study focused on the impact of digital contributors for the online user community.

1.1. Statement of the Problem

In the physical world, we know of everyday heroes who dedicate their lives to helping others. Their contributions to mankind inspire us to do good deeds and make the world a better place for future generations.

With the advent of technology, we now have individuals who are making a difference digitally by inspiring people and offering solutions to problems affecting humanity among other activities, contributing in the community and trying to make the digital activities better. Their dedication and contribution to the digital world could qualify them as digital heroes. Although in general, there is criteria to define someone as a hero and we have historical records for keeping heroes' history of contribution for mankind; there is no criteria set to define digital people as heroes. There is also no digital platform

[1] https://stackoverflow.com/about

to analyze digital people's activities in order to identify digital heroes and keep their records in one place.

This study therefore fills this gap by searching for the criteria to assist in defining someone as a digital hero. A web platform was also developed that can analyze the digital activities of online contributors, identify digital heroes and keep their records in one place. The hero selection criteria and the web platform would be the start of the journey of hero discovery to inspire digital people to help others accurately and frequently.

1.2. Aim of the Study

The aim of this research was to analyze the digital activities of users of the online developer community forum Stack Overflow and develop a web platform to discover digital heroes based on proposed selection criteria.

1.3. Objectives of the Study

The objectives of this research were to:

1) Establish a set of criteria to define a digital hero based on their digital activities.
2) Develop a conceptual model of the web platform for digital hero discovery.
3) Develop the web platform for hero discovery using the data mining approach.
4) Use the web platform to discover digital heroes on the Stack Overflow online forum.

1.4. Research Questions

This research was guided by the following questions:

1) What criteria set can be defined to identify digital heroes based on their digital activities?
2) What conceptual model will guide development of a web platform for digital hero discovery?
3) How can the data mining approach be used for developing a web platform to discover digital heroes?
4) What information is available in the Stack Overflow developer community online forums for use in discovering digital heroes?

1.5. Assumptions of the Study

This study assumed the following facts:

- The experts will provide their honest opinion about heroes' criteria based on their experience on the developer community forums.
- The opinions that will be collected will guide this study to finalize the selection criteria of digital heroes.
- Stack Overflow dataset will have enough relevant data that will help us to analyze and discover extra-ordinary contributors from the community.

1.6. Justification of the Study

In the society, heroes serve a purpose. They help people to get hope, encourage and provide examples for success. The findings of this study will provide motivation to developer community forum users to adapt good digital activities like helping other

people with their issues and secure the online activities by increasing good deeds due to intention of being a digital hero.

The developed web platform is useful not only to developer community but also to other domains in order to reward user contribution in the digital world. The findings will also be helpful in keeping the records of heroes in one place so that people including the future generations can get to know their contribution to mankind. The outcomes and recommendations of this study will serve as a reference model and data analysis process for further research for researchers interested in the area. Data process factors in the sense of hero discovery and digital hero selection criteria are still new in the digital world. Therefore, the findings of this study will add knowledge on defining and discovering digital heroes and reward people for their valuable contribution for the online community.

1.7. Scope of the Study

This study focused on analysis of people's digital activities in order to discover digital heroes from online community forums. The scope of the study was limited to a developer online community forum called Stack Overflow. The data of this study was limited to analyzing people's activities in reference to their contribution in community forums and developing a web platform to discover digital heroes.

1.8. Limitations of the Study

Discovering digital world heroes and defining someone as a digital hero is a rather new concept in the digital world. From the existing literature related to this area, not much has been written about it. The requirement that any scientific research requires a thorough literature review was therefore difficult to achieve. Due to this limitation, most

of the literature was obtained from the Internet and the work ensured that all the available of the related literature were reviewed.

There is no method of determining the accuracy of the personal information provided by the users on the community forum. Based on forum rules, only the person who asks the question can mark an answer as an accepted answer. The other people who view the answer cannot do the same. This cannot fully justify the answer as correct and highly useful. Another limitation is that most of the time the person who asks a question never come backs to vote an answer which solved their issue. As a result, there are good answers that do not get the votes they deserve.

1.9. Definition of Operational Terms

Web Platform: The term ‘web platform’ in this study is an online web-based platform which takes raw data as input, mines the data based on pre-defined variables, filters the data using hero selection criteria and provides a list of heroes as an output.

Open source: Refers to something that can be modified, shared and used for other purposes because it is designed to be publicly accessible. In general, open source refers to any program whose source code is made available for use or modified as users or other developers see fit.

Open source software: Open source software basically comes with source code that anyone can enhance, and modify. It is software which is made freely available and the original source code may be modified and redistributed.

Source code: "Source code" is the part of software code written by computer programmers that determines how a piece of software (program or application) works. Programmers who have access to a source code can improve and modify that program

by adding, updating or deleting features to it or fixing errors that cause problems to the software and to make sure that it always works correctly.

PHP: is a web-based scripting language and interpreter that is freely available. It stands for Hypertext Preprocessor.

HTML: Hypertext Markup Language is a text-based front-end web language that describes how content contained within a webpage is structured. The purpose of the markup is to tell a web browser how to display images, text, colors and other forms of multimedia on the webpage.

CSS: Cascading Style Sheets describes how HTML elements are to be displayed on a webpage. It is used to format the layout of webpages.

JavaScript: is a scripting programming language primarily used in web development. It is used to improve HTML pages and can be inserted anywhere within the HTML code of a webpage.

SQL: Structured Query Language (SQL) is a standard database language to manage relational databases and data manipulation. SQL is used to insert, delete, update and query data from the database.

MySQL: is a full-featured open source relational database management system (RDBMS) based on Structured Query Language (SQL).

1.10. Chapter Summary

This chapter covered a general introduction of this study, statement of problem, aim of this study, objectives, justification, scope and limitations of the study as well as definition of operational terms.

CHAPTER TWO - LITERATURE REVIEW

2.0 Introduction

The literature reviewed in this section is in-line with the objectives of the study and aims to get relevant information to answer the questions of this research. This chapter discussed earlier studies related to data mining, mining methods and related studies of data mining theories. It also discusses the conceptual framework, theoretical framework and other literature related to the study. This literature guided the researcher to mine information and to discover digital heroes based on their digital activities.

2.1.The Conceptual Framework of DHD Process

Digital Hero Discovery (DHD) web platform was conceptualized through the Knowledge Discovery in Databases (KDD) (Suba & Christopher, 2016; Fayyad, Piatetsky-shapiro, & Smyth, 1996) process of data mining. In order to discover new knowledge, this study used KDD process of data mining approach that works with dataset. The conceptual model of DHD data process model was designed based on the dataset and its data formation as shown in Figure-1.

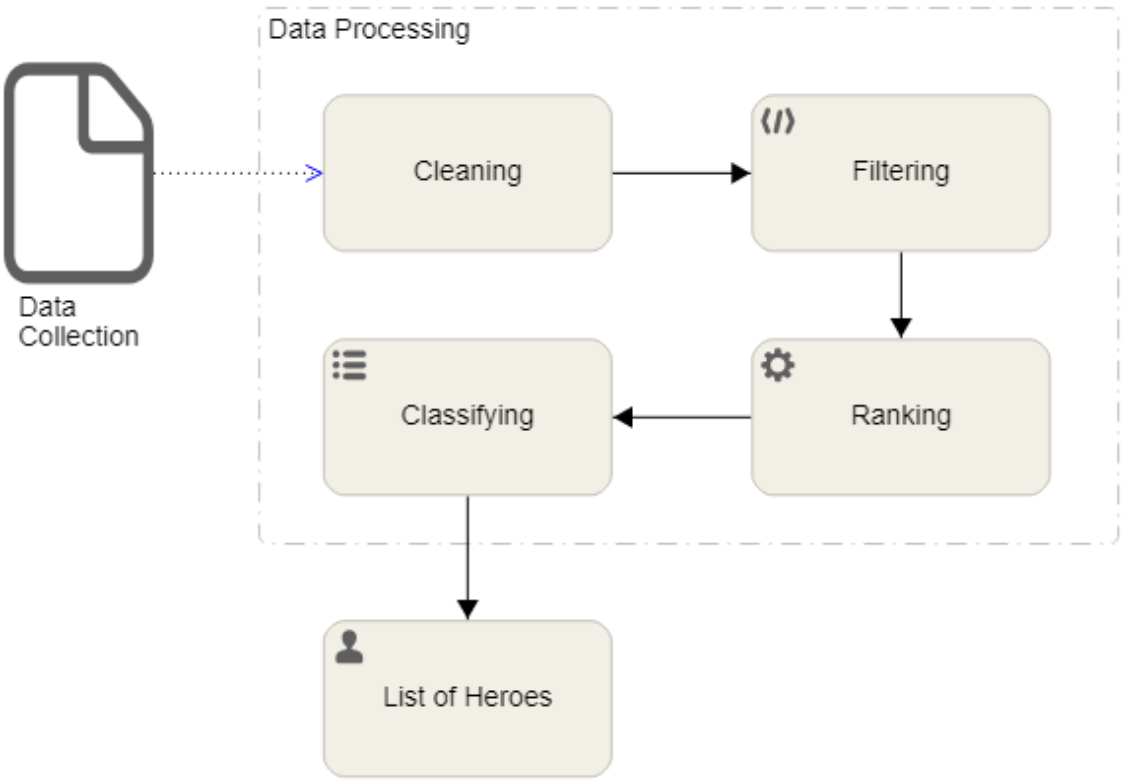

Figure 1: Data Processing Model of DHD Platform

The main elements of the DHD conceptual model are Data Collection, Cleaning, Filtering, Ranking, Classifying and Result (List of Heroes). The function of each element is described below:

Data Collection: DHD model requires initial data in order to process and produce the desired results. The work collected and used Stack Overflow dataset as an input to the DHD data processing model.

Data Cleaning: Data cleaning (also called data cleansing), is a process of detecting, removing or correcting incomplete, inaccurate, incorrect, irrelevant, corrupt, out-of-date, redundant, duplicate, incorrectly formatted, or inconsistent records from a record set, database or dataset. It cleans unnecessary data and removal of outliers or noise from data. There is often no established definition of the data cleansing, it varies depending

on the particular area in which the process is applied (Maletic, 2000). Several methods used to clean data include Sorted-Neighborhood Method, Selection of Keys, Equational theory and Computing the transitive closure over the results (Stolfo, 1998). Also, data could be cleaned by general methods including defining and determining types of error, identifying error instance and correcting the uncovered errors (Maletic, 2000). The DHD model used data cleaning process to gather only the necessary and relevant information from the inputted dataset that will be used in the next process.

Filtering: In the data processing system, filtering refers to merging, decoding, validating data using conditions and rules that are applied to the necessary process. Several algorithms used to filter relevant information exist. These include point based, segmentation based, and rule based algorithms (Sensing, 2000). The DHD conceptual model used rule based conditional algorithms to filter digital heroes based on selection criteria. Rule based conditional algorithm was used to apply conditions to each hero selection criteria as well as filter relevant information.

Ranking: In this study, ranking refers to scoring digital heroes based on their contribution in the community forum in order to acknowledge their activities of helping other people. Chapter 4 describes details about ranking of each hero selection criteria.

Classifying: Classification rule mining is one of the important data mining techniques. It aims to define a set of rules in the dataset or database in order to form an accurate classifier (Liu, 1998). Classification is learning a function that maps (classifies) a data item into one of several predefined classes. Classification methods are frequently used as part of knowledge discovery applications that also depend on applications purpose and help to identify objects automatically in large databases (Fayyad et al., 1996). The work used the classification rule to classify digital heroes based on their contribution ranks in order to differentiate one from another.

List of the Heroes: Finally, the DHD data processing model provides a list of digital heroes based on the input dataset and processing by each of the proposed steps of the model.

The researcher used this data processing model to analyze the Stack Overflow community members' data in order to discover digital heroes based on their community activities and hero selection criteria. This model could be useful to any other domain by changing its variables based on the domain and its dataset.

2.2. Theoretical Framework

Most researches in knowledge discovery and data mining are in databases and have focused on developing algorithms for numerous data mining tasks (Mannila, 2000). Recently some theoretical foundations of data mining have been proposed, also data is getting more complex and voluminous; researchers are working on developing better theoretical frameworks in the area of data mining (Muhammad, Mohamudally, & Babajee, 2013). This will help them to mine data in a more appropriate way as well as getting better outcomes from the data.

Some of the data mining theoretical frameworks that are well known and work better in data mining statistical resources are: Unified Data Mining Theory (UDMT), Probabilistic Approach, Data Compression Approach, Microeconomic View of Data Mining, Inductive Databases and Discriminant Analysis (Muhammad et al., 2013), (Mannila, 2000), (Fernandez, 2002).

In this study, Inductive Databases theory of data mining was used to achieve the objectives of the study. Inductive Databases use query concept to knowledge discovery and data mining (Mannila, 2000)

2.2.1. Inductive Databases Theory (IDT)

Relational database query use basic concept of data query as a powerful notion of query. The term inductive database refers to a normal database. In model-theoretic terms, the inductive database contains the data and the theory of the data (Mannila, 2000).

To mine related information from the dataset of Stack Overflow, this study followed inductive databases theory to query relationally the database to get specific and useful information from over 8.2 million users in the dataset. Because the best way to identify useful and relevant information from the Stack Overflow dataset is to use query and the result of the query also is inductive database, therefore it gives the closer property of relational databases.

2.3. Data Mining

Data mining is the process of examining data from different views and summarizing it into valuable and useful information (Sharma, 2016). According to Chitraa (2010), the automatic extraction of useful, unknown knowledge from an existing large database is called data mining. Using extract algorithms or factors filtering, it is possible to transform raw data to useful information. Data mining involves classification rules, searching patterns, clustering, association, statistical analysis and prediction analysis (Chauhan & Jaiswal, 2016). Cheung & Fu (2004) also stated that data mining approach is a group of methods to discover useful, legitimate, new and logical patterns in large dataset.

Concept of data mining is to clean, analyze and extract useful information or data from large, raw dataset and alter the information to a user readable structure for future use (Jayanthi, Kumar, Surendran, & Prathap, 2016). Data mining has been commonly used in different areas of information processing including stock market, banking, marketing,

education, healthcare, medicine, knowledge discovery, fraud detection, scientific discovery, prediction analysis and credit assessment (Chauhan & Jaiswal, 2016).

2.3.1. Data Mining Methods

A specific method may not be useful to all the applications due to appropriate data structure, application procedure and data variation of the method. Therefore the selection of a suitable data mining method depends on the purpose of the data process application and also on the data set compatibility (Chauhan & Jaiswal, 2016).

(i) Descriptive mining method

Descriptive data mining method discovers common factors or overall characteristics of the information from the database. The descriptive mining techniques include methods such as association, clustering, correlation analysis, feature extraction and sequential mining (Suba & Christopher, 2016).

(ii) Predictive data mining method

Predictive data mining method performs interpretation on input data to determine interesting and hidden useful information for future prediction. Predictive mining process includes classification, deviation and regression (Chauhan & Jaiswal, 2016).

In this study the predictive data mining method is the most suitable method for analysis of user information in order to identify digital heroes. Because predictive data mining method used classification as it’s way of discovering interesting information from the datasets.

2.3.2. Knowledge Discovery in Databases (KDD)

There are several data mining process models commonly used including Six Sigma data mining process model, CRISP-DM: Cross Industry Standard Process for Data Mining,

Scientific Data Mining Process Model, the SEMMA data mining process model, Data Mining Process Model, Hybrid Data Mining Process Model, and Knowledge Discovery Process (KDP) Model (Muhammad et al., 2013). In this study the researcher used KDD process in order to analyze community forum data and discover digital heroes. The study used KDD process because all its stages were fulfilling the data process model of DHD platform compared to other data mining processes. KDD guided to this study to develop data processing model of hero discovery.

Knowledge Discovery in Databases, refers to the process of discovering useful knowledge in a dataset, and highlights the "high-level" presentation of specific data mining methods (Fayyad et al., 1996). KDD is the acronym of datamining and it's important steps are data cleaning, data selection, data processing, data transformation, data mining, pattern recognition and knowledge recognition (Fayyad et al., 1996 & Jayanthi et al., 2017).

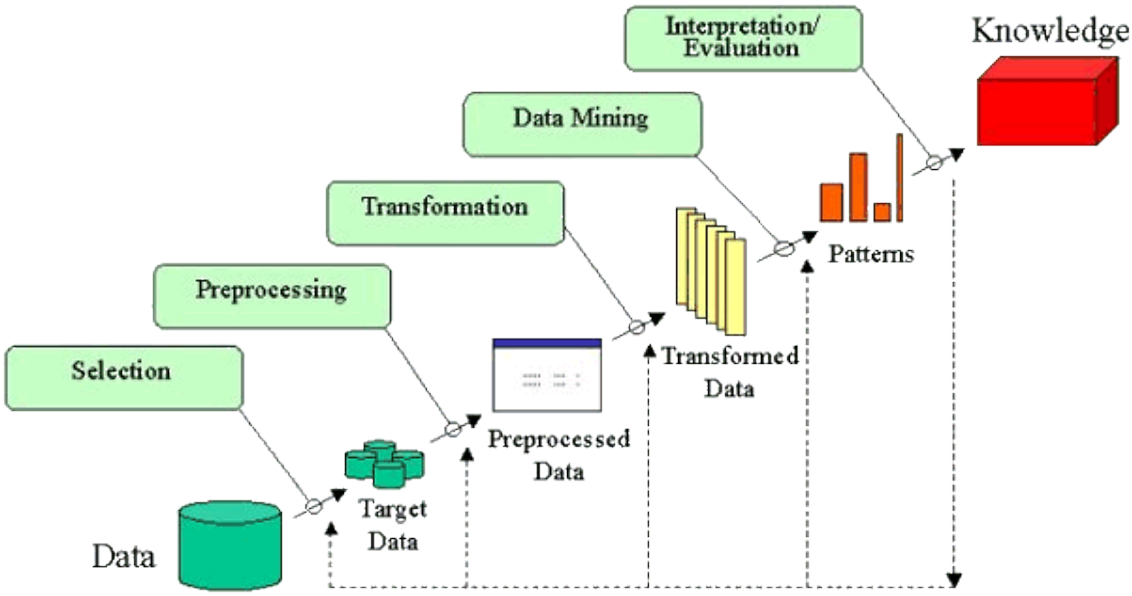

Figure 2: Steps that Compose the KDD (Fayyad et al., 1996)

The steps of the KDD process are as follows;

1) Understanding the domain: Obtain knowledge about the application domain and understand the KDD goals of the end-user.
2) Target Data: Selection of target data of importance from the data set or focusing on a subset of factors, on which discovery is to be achieved.
3) Data preprocessing: Cleaning the data by removing redundant, noisy, and irrelevant data and collecting necessary information to model.
4) Data reduction and transformation: Getting useful information from data set by selecting the feature and goal of the application. Using dimensionality transformation and reduce the effective number of factors under consideration.
5) Choosing the method of data mining: Identifying whether the goal of the KDD process is regression, clustering or classification based on the application purpose.
6) Patterns: Selecting the patterns of interest in a specific representational form as classification rules, clustering and regression.
7) Interpreting mined patterns.
8) Consolidating discovered knowledge.

2.4. Digital World

We are living in a digital world. According to Edward A. Fox (1995), the digital world that we are building can be called by many names, including Global Information, Cyberspace, Information (Super) Highway, Information Age, Paperless Society and Interspace. All of them are assisted by the Internet. The Internet, and World Wide Web (WWW), are recognized as universal means of connecting organizations and individuals (Pitta & Fowler, 2005). Nevertheless, their core is information. Over the network, there

are flows of information which are only accessible by electronic devices, operated by computers and saved in databases and libraries.

The digital world alters traditional conservation concepts from protecting the physical integrity of the object to specifying the creation and maintenance of the object whose intellectual integrity is its primary characteristic (Conway, 1996). Uzun (2015) defines digital world as a virtual environment that is built through computers and enhanced by the Internet and also it allows or contains storing and processing of digitized data.

2.5. Online Community Forums

An online community forum enables community members to learn, share, help others and enculturation of newcomers. It is also an opportunity to gain new insights from experienced experts into countless aspects of the practice (Gray, 2004). Online forums are made to get advantages of persistence, communications, accessibility and specificity. They are usually divided into topic based areas and all "threads" run based on the forum's focused area (Pitta & Fowler, 2005). These forums support communities designed around a specific interest. For example, developer community forums focus on programming related threads while scholars' community forums focus on research analysis related area. This study used a developer community forum to limit the study and discover digital heroes based on community activities.

2.5.1. Developer Community Forums

An online community forum is an Internet based platform where a group of people discuss their common interest by posting information and replying with their opinions. It refers to online discussion forums that provide an environment where members and customers come together and get peer-to-peer support. An online community combines

the capabilities of portals, forums and knowledge bases into one place (Dietz, 2016). In developer community forums people mostly discuss about programming related issues that they face in their work.

2.5.2. Question and Answer (Q/A) Forums

Online community forums content is mostly generated by the members' contributions. One of the relatively current appearances of this trend is question and answer (Q/A) sites. It is a place where users ask questions and others answer the questions (Harper & Raban, 2008). Question-Answer forums are rapidly becoming a valuable source of knowledge in many areas (Jurczyk & Agichtein, 2007). Q/A forums are the place where users request help and receive solutions for issues they face.

2.5.3. Stack Overflow

Stack Overflow is a Q/A web site designed to allow developer community users to ask their programming related questions and respond to the questions based on the topic. Programmers in the developers' community seem to have stopped reading programing or related books (Alex, 2014). When they want something and cannot figure it out, they just type a question into Google search engine. Sometimes, the first result from the search looks like it is going to give the exact question's answer that they are looking for and they are happy until they go to the link and find out it is a pay site or the answer is hidden behind a pay-wall. Sometimes the response might be wrong or not close to what they expected. To solve this problem Jeff Atwood and Joel Spolsky created Stack Overflow in 2008 (Alex, 2014).

Basically, all Q/A sites offer an interface designed for community activities like asking and answering issues (Harper & Raban, 2008). Usually, users will be asked an issue

with specific tags to categorize their issue, to route the question to get useful answerers. Likewise, most Q/A sites have the features of browsing and searching for a question, and filter by topics status including interesting, new, featured, weekly, monthly and tag basis. Stack Overflow has a highly featured interface (Figure-3).

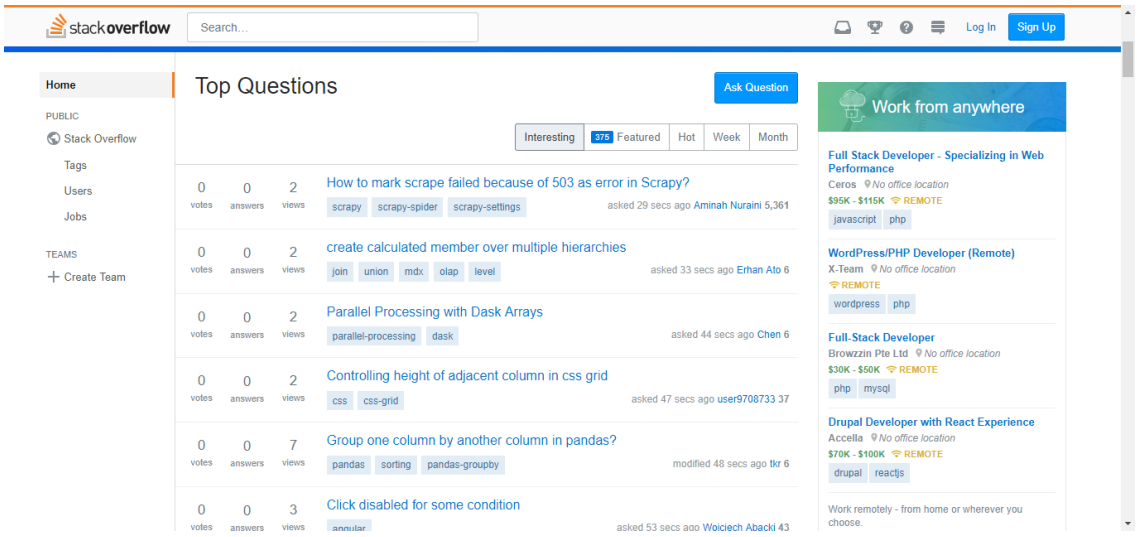

Figure 3: Stack Overflow User Interface (Screenshot: 09th July, 2018 at 10:35am)

Stack Overflow[2] is a popular online programing question and answer online community forum. The name for this forum website was chosen by using voting in April 2008 by the readers of Coding Horror, Atwood's popular programming blog (Slegers, 2015). As of 09th July 2018, average number of developers was 51,000, registered and unregistered users were 7.8 million, monthly visitors were over 50 million, questions over 14,000,000, answers over 19,000,000 and developers got help 7.5 billion times (Stack Overflow, 2018).

There are a number of question-answer community forums available on the web for the developer community. However, this study selected only Stack Overflow because it is the biggest developer community forum in the world based on popularity. Alex (2014)

[2] https://stackoverflow.com/company

wrote an article on CodeCondo.com[3] and compared 14 developer community forums by popularity as shown in Figure-4 (Cheung & Fu, 2004)

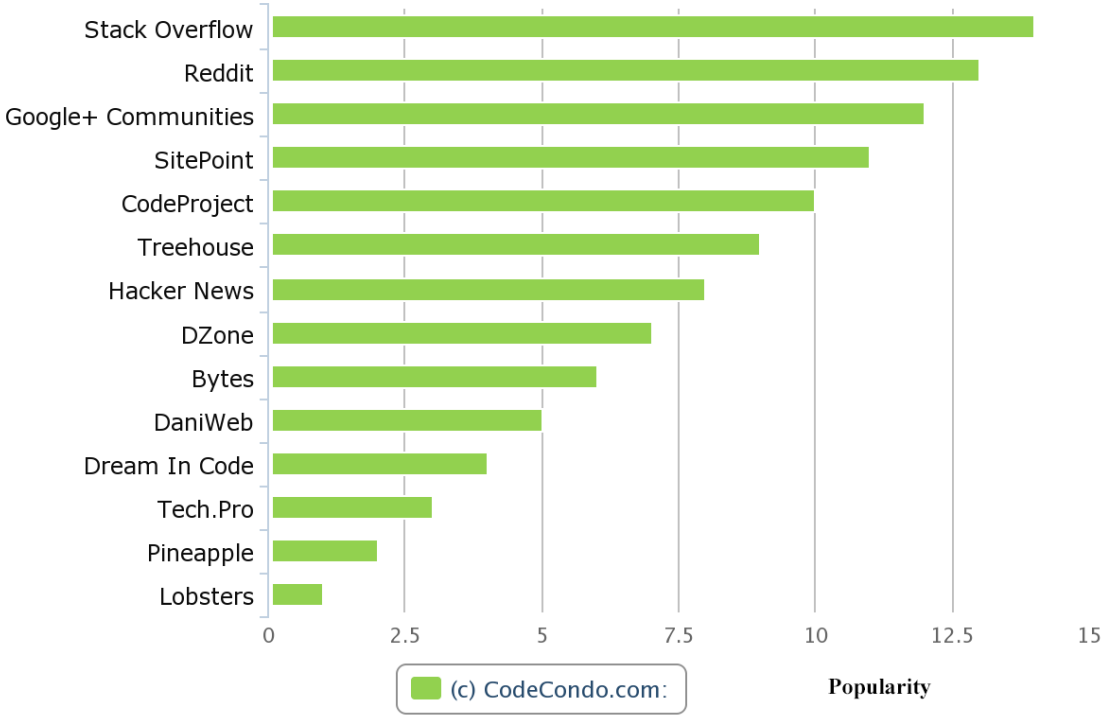

Figure 4: Programming Communities Based on Popularity – CodeCondo.com

2.6. Heroes

According to Miller (2011), heroes are people that usually go beyond and above in terms of the call of duty and do things that are extraordinary. They are people who save the day and have all the courage and responsibility. Furthermore, heroes give us models of good deeds that we aspire to do, whether those are our values that are related to integrity, honesty and courage (Miller, 2011). Zimbardo (2011) argues that we all have an inner hero. In addition he says, "The key to heroism is a concern for other people in need - a concern to defend a moral cause, knowing there is a personal risk, done without expectation of reward". McNally (2016) says anyone can be a hero and most of us are

[3] https://codecondo.com/programming-communities/

already on the journey of heroism. She points out that heroes change the world, take on tasks and save lives. In addition, she observes that each hero may have a unique cast of characters, a different quest and a specific setting but every hero's path is more or less the same.

2.7. General Qualities of Heroes

In general, there are some criteria to identify and declare someone as a hero. There are several processes to discover a hero based on specific criteria. These criteria used to find out whether someone is a hero are discussed in this section.

Courage and bravery come to mind whenever we think about heroism. It is truly difficult to achieve something heroic unless you are up against intimidating probabilities. As Nelson Mandela puts it, "Courage is not the absence of fear, but the triumph over it" (Murphy, 2014). Courage is the capability to challenge pain, danger and fear. A hero is recognized for his/her moral and physical courage. Whereas the moral courage is the ability to act morally in the right times of shame, discouragement and opposition, physical courage on the other hand demonstrate the brave act in times of hardship and pain. A heroic leader is courageous enough to take risks and is confident under pressure when others try to hide themselves (Tayyab, 2014). A hero continuously kills their fears and confronts any challenge head on; even the weakest drop of fear does not remain in the heart. Heroes show extraordinary courage essentially in battle or in disaster (Philip Zimbardo, 2011), (Republic of Rwanda, 2009).

A great American leader John F. Kennedy once said, “Do not ask what your country can do for you—ask what you can do for your country.” That is the true attitude of the heroic leaders. Heroes are more worried about group achievements than think about

their own goals (Tayyab, 2014). A heroic leader does everything with selflessness without any expectation of payback (Bill, 2014).

In the Republic of Rwanda (2009), they have a law for determining the decorations of Honor and National Orders for heroes. According to the law, a hero is any person who follows objectives and undertakes to get a special achievement for the public interest and with high proven sacrifice, integrity and noble courage in their acts and avoids being a coward in their actions. Based on their law, someone will be considered a hero, if they shall meet the criteria of proven integrity, patriotism, sacrifice, vision, proven courage or bravery, truthfulness, magnanimity and humanity.

"A National Hero is a person who is admired and acknowledged for their courage, outstanding achievements, and noble qualities; and is someone who has made significant positive contributions to the growth and development of society, and represents all of us" (Government of Bermuda, 2017). The Bermuda government also settled a law of selection process and hero criteria to determine national heroes. A national hero of Bermuda should meet most of the following criteria: they have made a significant and lasting contribution to Bermuda, has enriched the lives of others, his/her legacy stand the test of time and has relevance in the future, has contributed to the quality of life and destiny of Bermuda, has to be considered as outstanding in his/her area of activity, has a name recognized among the general population, has been recognized by professional body or organization, and has reflective knowledge of Bermuda's cultural heritage and diversity.

The Republic of Philippines (2015) settled a law, adopted by the Technical Committee of the National Heroes Committee on 3rd June, 1993 in Manila, to determine and identify someone as a national hero. Based on that law, a hero shall meet the following

criteria: have a concept of nation and thereafter aspire and struggle for the nation's freedom and contribute to a system or life of freedom and order for a nation, and contribute to the quality of life and destiny of a nation.

Additional Criteria for Heroes: (Adopted by the Technical Committee of the National Heroes Committee on November 15, 1995, Manila), a hero is part of the people's expression, thinks of the future, especially the future generations. The choice of a hero involves not only the recounting of an episode or events in history, but also the entire process that made this particular person a hero.

Though it is a slightly different criterion to identify and declare someone as a hero based on different countries government law, most of the criteria like courage, sacrifice, selflessness, extraordinary contribution for the country/society, working and thinking for future generations are basic criteria for defining someone as a hero.

CHAPTER THREE - RESEARCH METHODOLOGY

3.0 Introduction

This chapter describes the details about the research methodology in order to achieve the research objectives of this study. The study is quantitative in nature. Quantitative data was collected using questionnaires to design the hero selection criteria based on expert opinions. A dataset obtained from Stack Overflow was used to analyze user information based on the hero selection criteria to identify digital heroes. The methodology implemented in data mining, designing and developing the data processing model as well as web platform was Structured Systems Analysis and Design Methodology (SSADM).

This chapter includes research design, study population, study sample, sampling procedures, data collection, ethical considerations, data mining approach, theory building approach, data analysis, system analysis and systems design methodology.

3.1. Research Design

To achieve the stated objectives, this study used experimental research design. The research design of this study is experimental because of the need to measure digital activities of community people in the developer community forum in order to find out their contribution to the community. It was important that the design had to be appropriate for testing of the particular study hypothesis. According to Harper & Raban (2008), research that manipulates at least one independent variable, controlled some other applicable variables, also experiment of the outcome one or more dependent variables is called experiment research.

This research study uses a quantitative approach. A quantitative approach is one in which the investigatory primarily uses postpositive claim for developing knowledge (i.e., cause and effect thinking, reduction to specific variables and hypotheses and question, use of measurement and observation, and the testing of theories) employs strategies of inquiry such as experiment and survey and collect data on predetermined instrument that yield statistical data (Mark, Philip, & Tornhill, 2007).

According to Ary et al. (2010), "quantitative research is inquiry employing operational definitions to generate numeric data to answer predetermined hypotheses or questions". This study employed quantitative methods because all data related of this study including experts' opinion in the measurement of hero selection criteria is numerical and can be analyzed statistically.

3.2. Study Population

A population refers to a particular group of respondents who are of interest to the study and the researcher would like to generalize the outcomes of the study (Fraenkel and Wallen, 2003).

In order to develop digital hero selection criteria, the study also collected expert opinion about the possible parameters to be included in the selection criteria. The researcher targeted experts who work in the IT related field as developers as well as those who use Stack Overflow on a regular basis. The target population in this study was the 8,203,832 non-deleted registered users of Stack Overflow[4] developer community forum who registered their accounts between 2008 and 2017.

[4] https://stackexchange.com/sites

3.3. Sampling Procedures

Fraenkel and Wallen (2003) states that sampling is a method of choosing individuals who will participate in a research study. In this study, purposive sampling was used to select Stack Overflow users who contributed a minimum of one answer to any community question.

The characteristics of the Stack Overflow community users are the same in terms of their interests. All users have an interest in programming or software development. In that case, homogeneous sampling was used to gather a sample unit for the study. Homogeneous sampling is a purposive sampling technique that aims to achieve a homogeneous sample that shares the same (or very similar) characteristics (Mark et al., 2007).

Purposive sampling was used to identify a population of 537 experts from various organizations which includes 235 experts from universities, 282 from software companies, 10 from freelancers and 10 from other organizations. Convenience sampling was used to identify individuals who gave expert opinion on development of hero selection criteria. It is a non-probabilistic sampling technique that is used to select participants who are easy or convenient to approach (Fraenkel & Wallen, 2013).

Table-1 shows the study sample of 45 which constitutes 8.4% of the population. The study sample included 15 experts from university, 23 from software companies, 4 from freelancers and 3 from other organizations.

Table 1: Population and Sample size based on Organizations

Organizations		Country	Sample Size (S)
Universities	Moi University (MU) – Main Campus	Kenya	10
	National University of Science and Technology (NUST)	Zimbabwe	3
	University of Eastern Africa, Baraton	Kenya	2
	University of Illinois Urbana – Champaign	USA	2
	Technical University of Kenya (TU-K)	Kenya	2
Software Companies	XactIdea	Bangladesh	3
	Telenor Health	Bangladesh	5
	Pathao Inc.	Bangladesh	2
	Microlistics	South Africa	1
	Lets Learn Coding Ltd.	Bangladesh	2
	Pridesys IT Ltd	Bangladesh	5
	Logic Bits Solutions	Zimbabwe	2
	TheWebLab	Bangladesh	2
	CodePassenger	Bangladesh	1
	Business Automation Ltd.	Bangladesh	2
	SSL Wireless	Bangladesh	2
	ThemeBucket	Bangladesh	3
Freelancers			5
Other organizations			5

Population size: (P = 537) & Sample size: (S = 45)

3.4. Study Sample

The sample size of the study consisted of the 1,889,860 registered users of Stack Overflow who contributed a minimum of one answer to any post in the community forum. This study considers an answer or reply on a post as a contribution. Data of the forum users was mined using data mining approach and their digital activities analyzed with a view to discovering digital heroes.

Also, the targeted sample size in developing the selection criteria was the 45 experts from IT related fields. These are experts who have been using community forum for years to develop their skill as well as contributing to the forum. The study collected expert opinions from respondents to develop the hero selection criteria.

Table-2 shows that we have got 15 opinions from university, 23 from software companies, 4 from freelancers and 3 from other organizations.

Table 2: Respondents Based on Organizations

Organizations	Respondents (N)	Percentage (%)
University	15	33.3%
Software Company	23	51.1%
Freelancer	4	8.9%
Other organizations	3	6.7%

Number of respondents: (N = 45)

3.5. Data Collection

In order to achieve the first objective of this study, experts' opinions were collected using questionnaires to acquire relevant data for developing digital hero selection criteria. The criteria were used to collect and mine the relevant data from Stack Overflow community users.

As we discussed in Chapter 2, Stack Overflow is one of the most popular online community forums for the software developer community. It contains questions, answers and comments related with programming, algorithms and software tools to specific problems. To achieve objective number 4, we downloaded a dataset of all users since its launch in August 2008 to December 2017 in order to analyze their contribution and determine whether they are digital heroes based on the developed selection criteria. The user's information include user name, about, profile link, creation date, total up and

down votes, reputation score, number of posts, number of reply/answers etc. All this information from the dataset was used to analyze members' activities in order to discover digital heroes using the web platform. The dataset was then filtered using a query language and conditional rules according to the criteria set.

3.5.1. Questionnaires

According to Trueman (2015), a questionnaire is a sequence of questions that are asked to individuals to collect useful information about a specific topic. Questionnaires are commonly used in quantitative research. It is a valuable technique of gathering information from a great number of individuals, often referred to as respondents.

For the first objective, this study used questionnaires (see appendix 2) to collect the opinion from IT experts who are working in various organizations like universities, software companies as well as freelancers. In this particular case, structured questionnaires were used as recommended by Trueman, (2015).

3.6. Ethical Considerations

All Stack Overflow contents[5] are publicly accessible under the license of Creative Commons BY-SA 3.0 license[6] including questions, answers and user profile information. They provide 3 ways to collect and use information for any tool or research purpose. The ways are Stack Exchange Data Dump[7] that are updated every 3 months, using Stack Exchange API[8] and Stack Exchange Data Explorer[9] which use direct query in their database. This study used publicly accessible data provided by Stack Overflow and the researcher abided by their terms and policy.

[5] https://stackoverflow.blog/2014/01/23/stack-exchange-cc-data-now-hosted-by-the-internet-archive/
[6] http://creativecommons.org/licenses/by-sa/3.0/
[7] https://archive.org/details/stackexchange
[8] https://api.stackexchange.com/
[9] https://data.stackexchange.com/stackoverflow/queries

3.7. Data Mining Approach

For answering research question number 3, how can the data mining approach be used for developing a web platform to discover digital heroes?, the study used data mining approach to mine Stack Overflow users' information in order to discover digital heroes. Knowledge Discovery in Databases (KDD) process was followed to develop a DHD data processing model that helped to mine data and classify discovered users as digital heroes based on Stack Overflow community forum dataset.

3.8. Research Phases

This study was divided into three phases as shown in Figure-5. The first phase was the establishment of the hero selection criteria and collecting experts' opinion to get measurable values for the criteria. The second phase involved collecting information from Stack Overflow dataset that is relevant to the study. The third phase entailed developing a data processing model using data mining approach. It was also used as a guideline for developing the web platform for analyzing and discovering digital heroes based on Stack Overflow community members' contributions.

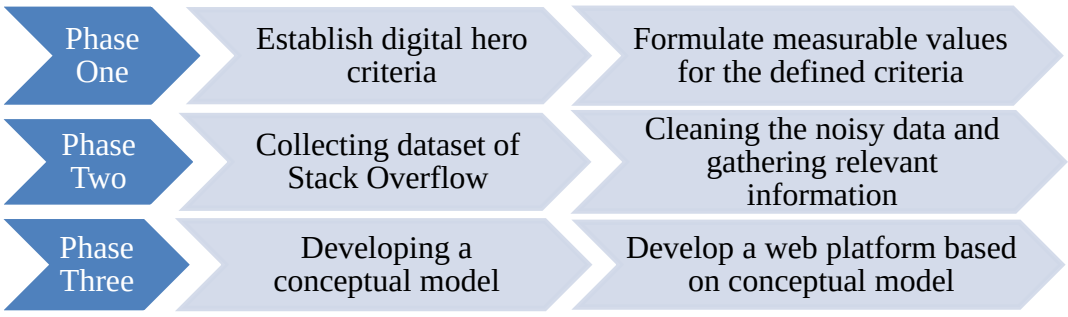

Figure 5: Research Phases to Achieve Research Objectives

3.9. DHD Platform Testing

The web platform to discover digital heroes based on digital activities using data mining approach was tested to validate its process of data mining and analyzing of relevant data to meet the valid outcomes of the system. The researcher used 1000 Stack Overflow users' information to test the DHD Platform before using all users' information from the dataset.

3.10. Data Analysis and Presentation

The collected information was input to the DHD web platform. The data analysis was done as a system background process. Outcomes of the system were presented using the web interface of the DHD web platform. All data were analyzed in order to discover the digital heroes based on their digital activities using the data mining approach. Contributor activities were analyzed based on defined selection criteria for the digital heroes. The platform also provides a proposed score in order to differentiate the level of one hero from the other based on their online contribution and their records are stored in the DHD platform database. The platform also provides a search system where digital heroes can be searched by user's name, country and hero category.

3.11. System Analysis, Design and Methodology

After the digital hero selection criteria was developed and dataset collected from Stack Overflow completed, the Digital Hero Discovery (DHD) web platform was used to mine all provided data of the community users, discover qualified users who meet the hero selection criteria, rank them based on their activities and categorize them based on rank. This web platform collects data from Stack Overflow developer community forum. To achieve research objective number 3, the following steps were followed:

- A web based platform was developed based on the proposed conceptual model to analyze user activities in the developer community forum.
- The platform was developed by using PHP web programming language, hypertext markup language (HTML), MySQL database, cascade style sheet (CSS), JavaScript scripting language and Structured Query Language (SQL).
- The platform provided as an output a list of digital heroes based on their online activities and selection criteria.

The final outcome constituting a list of digital heroes was stored in the database and the records displayed on the web platform for future generations.

3.11.1. Systems Methodology

In system development, many methodologies exist and each is appropriate for a specific type of application. There are many methodologies that are used for developing computer-based information systems. These methodologies include Systems Development Life Cycle (SDLC), System Prototype Method (SPM) and Structured Systems Analysis and Design Methodology (SSADM) (Saleemi, 2007).

In this study, SSADM method was chosen because it is a widely-used computer-based information system development method. SSADM distributes an application development project into stages, modules, tasks and steps (Margaret, 2008). It was developed by Learmonth Burchett Management Systems (LBMS) and the Central Computer Telecommunications Agency (CCTA) in 1980-1981 as a standard for developing British database projects (Techopedia.com, n.d.).

3.11.2. SSADM Stages

Select Business Solutions Inc. (2018), explain that the SSADM method includes the sequence of an application analysis, design tasks and documentation concerned with:

- **Feasibility Stage**: high level analysis of the current situation. A Data Flow Diagram (DFD) is used to describe the DHD platform background data processing system working procedure and to visualize recognized problems. This stage involved developing an activity model by investigating the current requirements, processing and data.

- **Requirements Analysis**: it involves circulating questionnaires, interviewing employees and observations. It is important to understand the system requirements as it is the starting part of the project. The study analyzed all requirements of the DHD platform by following the conceptual model before starting to code the system (see Section 5.3).

- **Requirements Specification**: it is the most complex stage. This stage requires developing a full logical specification of the new system requirements and the specification must be free from ambiguity, error, and inconsistency. To develop DHD platform system specification we followed sever side and client requirements of the system (see Section 5.3.1).

- **Technical System Specifications**: as the system specifications, this stage entails the choice of technically feasible options that will determine implementation. In this stage of DHD platform the researcher analyzed the specification of implementing the system to another domain too (see Section 5.4).

- **Logical Design**: this stage specifies the key methods of interaction in terms of command structures and menu structures. In the DHD platform the researcher designed the input and output of the system (see Section 5.5).

- **Physical Design**: it is the last stage of SSADM. In this stage all the logical specifications of the system are converted into descriptions of the system in terms of software and real hardware. In terms of database structures, the logical data structure is converted into a physical architecture. The DHD platform physical stage covered database design, Entity Relationship Diagram (ERD) and entities (see Section 5.6).

SSADM uses the top-down approach. It adopts a waterfall method where each phase is approved and completed before the subsequent phases can be initiated (Select Business Solutions Inc., 2018).

3.12. Chapter Summary

This chapter discussed the research methodology that this study adopted. It explained how the sample was determined and the data collection procedures carried out. It further explained the systems analysis and design methodology adopted in developing the digital hero discovery web platform.

CHAPTER FOUR - DATA PRESENTATION, ANALYSIS AND INTERPRETATION

4.0 Introduction

This chapter presents the analysis of data to achieve the objectives of the study. The information obtained through completed questionnaires and dataset provided the basis for data presentation, analysis and interpretation. Analysis was guided by the research objectives stated in chapter one.

4.1. Establishment of Digital Hero Selection Criteria

To establish digital hero selection criteria as stated in research objective one, establish a set of criteria to define a digital hero based on their digital activities, the study was guided by the existing literature and expert opinions. The literature had clear criteria for identifying general life heroes. However, to develop digital hero selection criteria based on the digital activities, the work used questionnaires to collect expert opinions to come up with parameters for digital hero selection criteria measurements.

Using questionnaires, the study sought for relevant information from 45 respondents in order to establish digital hero criteria with measurable value. A total of 28 out of 45 respondents returned the questionnaires. The respondents who returned the questionnaires included university (10), software companies (13), freelancers (2) and other organizations (3). The parameters for the criteria included experience in the community forum, accuracy in their digital activities, continuous activity in the community, and trust by the community members.

4.1.1. Experience

Experience in the community and community activity are very important factors for anyone to be considered a digital hero. Experience in their specific areas of expertise is equally crucial. In the community forum, a digital hero should respect forum terms and policy and other members as well as support community members when they need help based on the hero's skill and experience in the related area.

The respondents were requested to give their expert opinion on the number of years a digital hero should have worked on community forums. Table-3 shows their responses.

Table 3: Hero Criteria of Experience in the Stack Overflow Community Forum

Experience	Respondents (N)	Percentage (%)
1 to 5 years	12	42.9%
6 to 10 years	**14**	**50%**
11 to 15 years	1	3.6%
16 to 20 years	1	3.6%
More than 20 years	0	0%

Number of respondents: (N = 28)

The opinion of most of the respondents in Table-3 showed that for someone to be considered a digital hero they should have worked on community forums for an average of 6 to 10 years. This was suggested by 50% of the respondents. A high percentage of 42.9% of the respondents gave an opinion of 1 to 5 years' experience. However, there were few respondents with the opinion that a digital hero should have worked for more than 10 years with none of the respondents suggesting experience of more than 20 years. In general, over 57.2% of experts agreed that a digital hero should have more than 6 years of experience in the community forum. These results show that although experience is important for someone to be considered a hero, they do not necessarily

need to have many years of experience. People improve their skills and knowledge to perform like those who are skilled, and more experienced (Kinsella, Ritchie, & Igou, 2016).

4.1.2. Accuracy

Based on the digital world, digital heroes should be accurate in their digital activities. Any activities like answers or replying to a post or providing help in someone's issues must be accurate and accepted by the community.

The respondents were asked to comment on the accuracy of information that a digital hero should provide in community forums. Table-4 shows their perception on different ranges of accuracy.

Table 4: Hero Criteria of Accuracy

Accuracy	Respondents (N)	Percentage (%)
0 to 50%	0	0%
51 to 60%	2	7.1%
61 to 70%	5	17.9%
71 to 80%	8	28.6%
More than 80%	**13**	**46.4%**

Number of respondents: (N = 28)

Table-4 shows that a majority of the respondents indicated that digital heroes should have more than 80% accuracy in their activities. All the respondents indicated that for one to be considered a digital hero they should contribute information that is at least 50% accurate with the highest required level of accuracy ranging from 71 to 80% and from 80% to 100%. Accuracy of information is a very important aspect to all individuals. It increases trust and reliability of the person giving the information. Heroes

should be reliable individuals in any community forum because based on their contribution somebody can make a decision to modify some part of their forum (Eslake, 2006).

4.1.3. Activeness

For someone to be considered as a digital hero, he/she needs to be active for specific time duration in the community forum. If they stop their activity for a long time, they will be counted as less active members of the community forum.

The respondents were requested to give their opinion on the length of the continuous period of activity a digital hero should have participated actively in the community forum. Table-5 shows the results of their expert opinion.

Table 5: Hero Criteria of Activeness

Activeness	Respondents (N)	Percentage (%)
1 year	4	14.3%
2 to 3 years	5	17.9%
3 to 5 years	**12**	**42.9%**
6 to 10 years	5	17.9%
10 to 20 years	2	7.1%
More than 20 years	0	0%

Number of respondents: (N = 28)

In Table-5, majority of experts (42.9%) indicated that digital heroes should demonstrate 3 to 5 years of continuous activeness in the forums. In general therefore, experts believe a digital hero should have a minimum of 3 years of continuous activeness in the online community forum. Later this study considered to find out how many answers provided by a user each month in 3 years to differentiate most active user to the community

forum (see Section 4.2.3). However, it is not necessary for the heroes to have been active for more than 20 years as indicated by 0% of the respondents. For anyone to be considered a real life hero they should show their concern for the people and be willing to dedicate much of their time to helping others. They should be interested in assisting people and providing solutions to their problems to encourage others and strengthen the community (Laeeka Khan, 2017). These findings show that the same applies for digital heroes; they should be people who are willing to continuously assist online community members and provide solutions to their problems.

4.1.4. Trust

Digital heroes need to be trusted by the community people. If they are doing selfless activities and community members are getting help and support from them, they will be trusted by community users. In life in general, people follow those that they trust (Eslake, 2006), therefore to determine the level of trust of the online community members their reputation score of Stack Overflow website can be studied. Reputation[10] score is a rough measurement of trust that can be earned by the Stack Overflow community members.

The respondents were asked to give their views on the minimum number of reputation score of a community member to be trusted in the community. Table-6 shows the responses.

[10] https://stackoverflow.com/help/whats-reputation

Table 6: Hero Criteria of Trust

Reputation Score	Respondents (N)	Percentage (%)
10,000 – 19,999	2	7.1%
20,000 – 49,999	**14**	**50.0%**
50,000 – 99,999	5	17.9%
100,000 – 499,999	3	10.7%
Over 500,000	4	14.3%

Number of respondents: (N = 28)

Table-6 shows that for someone to be considered a digital hero it is important for them to have a good number of reputation score with only 7.1% of the respondents suggesting that they need less than 20,000 reputation score. A total of 92.9% of experts were of the opinion that a digital hero should have more than 20,000 reputation score with the highest percentage (50.0%) saying that a digital hero should have at least 20,000 reputation score. Over time community members trust one another based on their community activities and it increases the value of information source (Pitta & Fowler, 2005).

Stack Overflow also have policy of trusting community members in order to provide community privilege levels that allow use of some tools that are available on the site. A member with a minimum number of 20,000 reputation[11] score is considered as a trusted user in the community. In this study, any member who crossed the 20,000 reputation score as trusted by the community for the sake of fulfilling hero selection criteria of trust.

[11] https://stackoverflow.com/help/privileges/trusted-user

4.1.5. Summary of Responses

According to this study experience is the number of years that someone has spent in the community forum doing various activities including contributing to the forum, commenting on other people issues, voting on other contributor's contribution, posting new articles, joining a discussion group, raising new topics and making corrections. Accuracy refers to correct information that an individual member contributes to the community. For example, answering someone else's problem and people voting that they are actually getting help from the answer because it is accurate and useful. Activeness in this study means that a community member has to do continuous activity for a minimum period of time. In this study experts agreed on a minimum of 3 years of continuous activity such as commenting, answering on other's issues, replying to other people questions, posting new thoughts. In community forums people get trusted because of their useful contributions. People vote for their contributions and forums also provide ratings based on the contributions they have done. These criteria are used in considering someone as a digital hero based on digital world activities.

From the results we propose the developed hero selection criteria for identifying digital heroes as shown in Table-7.

Table 7: Hero Selection Criteria for Identifying Digital Heroes

Criteria	Value	Respondents (%)
Experience	At least 6 years	57.2%
Accuracy	At least 80%	46.4%
Activeness	At least 3 years	82.2%
Trust	More than 20,000 reputation	50.0%

Based on expert opinion, this study found out that someone could be considered as a digital hero if he/she has more than 6 years of experience, more than 80% of accuracy ratio, more than 3 years of continuous activeness and trusted by reputation score more than 20,000 as shown in Table-7.

4.2. Proposed Scoring Procedure

The study developed digital hero selection criteria and its numeric measurement in order to analyze data using data mining approach. It is obvious that a few Stack Overflow community people would qualify as digital heroes based on the selection criteria but it is also important to differentiate one hero from the other based on their activities. For example, a user could pass the criteria of 6 years of experience in the community, but some other user could have 7 years or 8 years of experience. On the other hand, a user could have 80% of accuracy requirement in order to fulfil hero criteria but another user could have 90% of accuracy. The same applies for other criteria. To solve this issue, the study proposed a scoring procedure that also guided the data mining process of the hero discovery web platform. There are no standard ways of scoring or grading system for hero selection criteria. Studies have proposed scoring or grading system based on their particular studies and it fits only their area of study and purpose of scoring. This study therefore proposed a scoring procedure for each selection criteria parameter. The proposed scoring procedure for this study helps to differentiate one user score from another and appreciate their level of contribution in the digital community forums.

4.2.1. Experience Scoring

In order to categorize users based on their experience in the Stack Overflow community forum, for users who fulfilled the criteria of 6 or more years of experience, the study proposed a procedure of scoring shown in Table-8.

Table 8: Scoring Procedure of Experience Criteria

Years of experience in the community	Exp. Score
10	5
9	4
8	3
7	2
6	1

As Stack Overflow community forum was created 10 years ago in the year 2008, the study considered 10 years as the highest level of experience. Also, a digital hero should have at least 6 years of experience in the community in order to fulfill selection criteria of experience. Table-8 shows that user who has 6 years of experience in the community will get 1 point score and it increases by 1 for each year based on their years of experience in the community. A user with 10 years of experience will have a score of 5.

4.2.2. Accuracy Rate and Scoring

Most community forums have the feature of voting in a post. Users can give a positive vote (up vote) or negative vote (down vote). Stack Overflow has a feature of voting in the question or answer in order to check the quality of the questions or answers[12]. The study collected these votes (up and down votes) from the dataset to calculate the accuracy of each member's activities.

In order to calculate accuracy rate of a user's contribution in the Stack Overflow community forum, the study considered, up and down votes and calculated accuracy using Equation 4.1.

[12] https://stackoverflow.com/help/why-vote

$$\text{Accuracy Rate} = \frac{Total\ number\ of\ Up\ votes * 100}{Total\ Votes\ (Up+Down\ votes)} \quad \text{Equation: 4.1}$$

Example:

From the dataset there is a user who has total number of 43,953 up votes, 254 down votes and total votes of 44,207 in all provided answers.

$$\text{Accuracy Rate} = \frac{43{,}953 * 100}{44{,}207} = 99.43\%$$

After the calculation by proposed equitation the accuracy rate is 99.43%. In order to fulfill digital hero selection criteria, a user requires more than 80% of accuracy rate.

To differentiate one user from another user based on their accuracy rate, this study proposed the accuracy scoring procedure as shown in Table-9.

Table 9: Scoring Procedure of Accuracy Criteria

Accuracy rate range	Acc. Score
97-100%	5
93-96%	4
88-92%	3
84-87%	2
80-83%	1

As it is stated, a digital hero activity should be at least 80% accurate and maximum accuracy is 100%. We divided this range (80% - 100%) into 5 in order to keep the point score range within 1 to 5 as shown in Table-9. These range and point scores guided us in the data processing model as described in chapter 5.

4.2.3. Activeness Scoring

Based on expert's opinion on activeness criteria as a requirement to be considered as a digital hero, a user should be continuously active for a minimum of 3 years. In order to differentiate user activeness in the community forum, this study considered each month contribution done by the user. In the dataset, the study found 5,600 users who contributed a minimum of one answer in the community. It also collected a total number of answers contributed by users in 3 years. This study divided total number of contributions by 36 to get each month's contributions. After dividing by 36 months (3 years), the study found in the dataset that the highest number of answers that were provided by a user in a month is 681.14. It considered a maximum number of 700 answers and divided them by 5 in order to get a 5 range set of 140 answers in each set. They then proposed the following scoring procedure (Table-10) to differentiate users based on their contributions in the community.

Table 10: Scoring Procedure of Activeness Criteria

Answers range (per month) in 3 years	Act. Score
561-700	5
421-560	4
281-420	3
141-280	2
1-140	1

Table-10 shows that an average of every 140 answers that were provided by a user in a month within 3 years will receive 1 point.

4.2.4. Trust Scoring

It is stated that a digital hero should be trusted by the community. Based on Stack Overflow, the study consider a reputation score[13] that is provided based on the user's various activities including asking questions, answering questions, up and down votes. Stack Overflow also considers a user as a trusted user[14] when a user gets 20,000 of reputation score (see section 4.1.4). The work collected this reputation score from the dataset and proposed a scoring procedure that is shown in Table-11 in order to differentiate users from one to another based on their reputation score.

Table 11: Scoring Procedure of Trust Criteria

Reputation range	T. Score
722,001 – 800,000	5
644,001 – 722,000	4.5
566,001 – 644,000	4
488,001 – 566,000	3.5
410,001 – 488,000	3
332,001 – 410,000	2.5
254,001 – 332,000	2
176,001 – 254,000	1.5
98,001 – 176,000	1
20,000 – 98,000	0.5

In the dataset the maximum reputation score was 790,065 of a user who qualified in all criteria. As reputation is a big scale of number and in order to keep the scoring point within 1 to 5, the study considered 800,000 as a maximum level of reputation number and divided them into 10 sets of range by considering each range for 0.5 points as shown in Table-11.

[13] https://stackoverflow.com/help/whats-reputation
[14] https://stackoverflow.com/help/privileges/trusted-user

4.2.5. Digital Hero Scoring

After considering scores for all criteria, the study calculated a final score for a digital hero by using average formula as shown in Equation 4.2.

$$\text{Hero Score} = \frac{Exp.Score + Acc.Score + Act.Score + T.Score}{Total\ number\ of\ criteria} \qquad \text{Equation: 4.2}$$

Where:

Exp. Score = Experience Score

Acc. Score = Accuracy Score

Act. Score = Activity Score

T. Score = Trust Score

Example:

For example, a user has over 8 years of experience in the community and thus earned 4 experience score points. The user accuracy rate was 97.36% and that earned them 5 accuracy score points. The average number of posted answers per month was 185 so the user earned 2 activity score points. Reputation score in the community was 278,763 and earned the user 2 trust score points. Using Equation 4.2, the hero score points in all four criteria was summed up and divided by 4, shown as follows:

$$\text{Hero Score} = \frac{4 + 5 + 2 + 2}{4} = \frac{13}{4} = 3.25$$

After calculating all score points that were gained by qualified users in all four criteria, this user earned hero score of 3.25, which was classified using the proposed hero classification procedure as described in next section.

4.2.6. Digital Hero Classification

Classification is one of KDD process of data mining approach. This study used data mining approach in order to discover digital heroes by analyzing dataset of Stack Overflow community forum member's activities in the community.

Based on hero score, the study proposed the following classification (Table-12) to classify digital heroes that were discovered after data processing and hero selection criteria.

Table 12: Scoring Procedure of Digital Hero Classification

Hero Score	Category
>= 4	A
3 – 3.99	B
2 – 2.99	C
< 2	D

Hero classification is to categorize digital heroes based on their earned hero score in proposed four categories: A, B, C and D based on their contributions in the community, where category A represents the hero with the highest contribution to the community. Table-12 shows that any qualified user who earned a hero score of 4 and more was considered as an A-category heroes, where users who earned hero score between 3 and 3.99 were considered as B-category heroes, also users whose hero score was within 2 and 2.99 was considered as a C-category hero and any users who earned less than 2 hero score were considered as D-category heroes.

4.2.7. Data Range Modification

This study used a dataset of Stack Overflow users between 2008 and 2017. As discussed in previous section all criteria scoring range was considered based on the maximum number that was found in the dataset. For example, the study considered the maximum number of the reputation score as 800,000 in order to set ranges and score them within the point of 1 to 5, because in the dataset the maximum reputation score was 790,065 of a user who qualified in all criteria. All these considerations of maximum values could be changeable based on the dataset in order to fit them within the score range of 1 to 5.

4.3. Chapter Summary

This chapter discussed data analysis and data presentation of this study. It explains how the study developed hero selection criteria and what the responses were. It further explains the proposed scoring procedure of the digital hero for each hero selection criteria.

CHAPTER FIVE - DEVELOPMENT OF THE DIGITAL HERO DISCOVERY (DHD) PLATFORM

5.0 Introduction

DHD is a web platform that runs on the web and depends on the Internet and is completely separate from a computer operating system. It requires a browser and Internet to connect to the server in order to view what is in that platform. The aim of the study was to develop a web platform to discover digital heroes based on their digital activities for the sake of helping the digital community people. The platform is an open source system that analyzes user's information and discovers digital heroes based on selection criteria. It was developed to work with Stack Overflow community users' activity data. DHD platform is written in PHP web scripting language that is installed in server side, along with MySQL database and Apache web server. In this chapter design and development of the Digital Hero Discovery web platform is discussed.

5.1. Systems Analysis and Design Methodology

As discussed in section 3.13.1, Structured Systems Analysis and Design Methodology (SSADM) was used in the design and development of the DHD web platform. SSADM is a data driven, 'waterfall' systems approach to the analysis and design of information systems. SSADM follows the waterfall life cycle from the feasibility study to the physical design stage of system development as elaborated in section 3.13.2.

5.2. Systems Analysis

Systems analysis entails learning about the current systems, making recommendations, finding alternative solutions to solve problems, detailed documentation and its cost. In addition, systems analysis also explains why a system is done the way it is and determines improvements and changes. Therefore, appropriate systems analysis was

carried out to investigate the important aspects which are needed to make the DHD Platform system workable and get correct outcomes from it. To mine the dataset, this system follows several stages of DHD data processing conceptual model such as data cleaning, data filtering, ranking and classification. The conceptual model guided the development of specifications for the DHD platform.

5.2.1. Investigation of the Current Existing Systems

Best knowledge of this work, there are no existing data mining systems that can analyze digital people's contributions and discover someone as a digital hero based on selection criteria and mining process. DHD web platform is a system that analyzes inputted data of thousands of Stack Overflow users and their contributions. In order to identify extraordinary users based on their digital contribution it uses the KDD process of data mining approach and digital hero selection criteria.

5.2.2. Benefits of the DHD Web Platform

The DHD web platform will provide a new way to identify and recognize people's contributions and keep historical records of digital heroes in one place just like the general life heroes. It will help future generations to know about community members' contributions and increase good deeds for the intention of being a digital hero. The DHD platform will provide some more benefits which include:

- It is an open source, free and easy to use system.
- Visitors can learn about digital heroes and their contributions.
- It is easy to search a listed hero by name, hero category or country name.
- Community users can learn the digital hero selection criteria, hero category and required level of contribution for the sake of being a hero.

- It is an online web platform that can be accessed anywhere, anytime through the Internet.
- The model and the platform can be extended to analyze different type of domain by changing criteria set and data mining variables based on domain type and its data type.

5.2.3. Inputs to the DHD Platform

DHD platform used KDD process of data mining approach in order to mine Stack Overflow community members' information in order to identify digital hero based on their digital contributions for the community. The DHD platform mined only relevant information for the system that was considered to discover digital heroes including questions, answers, comments, user logs, flag information and delete records. The user's information including user name, profile picture, data of account creation, community reputation score, total number of answers, total number of positive votes (up votes), and total number of negative votes (down votes).

5.2.4. Expected Processes in the DHD Platform

As the DHD conceptual model, there are a number of processes that took place in order to analyze users' information based on the selection criteria, rank them based on the proposed ranking procedure, and classify them based on the proposed category (see chapter 4). The process the DHD web platform followed is as below:

- Took raw dataset as input to the system
- Cleaned the data in order to separate only the useful information that is relevant to the process.
- Filtered the cleaned data based on digital hero selection criteria set.

- Ranked the community users based on proposed ranking procedure in order to differentiate one user from another.
- Classified the community users based on the proposed digital hero category and their rank.
- Finally, the system provided a list of digital heroes and displayed them in the web platform interface.

5.2.5. Expected Outputs from the DHD Platform

As discussed in the previous section, the DHD web platform provides a list of digital heroes after analyzing their digital activities in the community forum. Stack Overflow developer community forum was used to acquire user's information. The platform displays user profile with their contribution records in a nice user interface and it also provides an easy search form to find out any hero by name, hero category as well as by country. It has a Report navigation which helps to know a digital hero based on their activity in the community, graphs and tables based on top contribution area, top hero by country, by year etc.

5.3. Requirements Analysis and Specifications

As mentioned before, the DHD platform is a web-based platform. The study used a web hosting platform to host the DHD platform and make it easily accessible to any user by using Internet connection. In the hosting platform, there are software requirements that are required in order to run the DHD platform.

5.3.1. Recommended Software Requirements

To implement DHD platform so that it can be accessed globally, we recommend some tools and software as below:

Server Software:

- **Operating System**: Windows or Linux operating system to host DHD platform
- **Database and Web Server**: DHD platform requires a database system and web server to facilitate access of stored data and its output.
    - **Database**: To store analyzed data this study used MySQL[15] database management system. It is one of the standard query languages for interacting with databases. MySQL is an open source database server that is fast, reliable and free. It is also cross platform supported and provides high performance.
    - **PHP**[16]: stands for Hypertext Preprocessor. It is one of the popular scripting languages that are especially suited to web development. It is flexible, fast and pragmatic. Most of the popular blog sites, websites and web applications in the world were developed using PHP language.
    - **Apache**[17]: is one of the most commonly used and popular web servers that are available for free. Apache was chosen because of its suitability as a web server. It is part of the Apache Software Foundation.

Client software:

- Windows or Linux operating system for laptop or desktop computer
- Android or iOS operating system for tablet or mobile phone.
- Any web browsers such as Mozilla Firefox, Chrome, Safari, Opera etc.
- Internet connection

5.4. Systems Customization and Implementation

In order to use DHD platform, it is important to set the digital hero selection criteria (see Section 4.1) and data processing variables based on the domain type and its data.

[15] https://www.mysql.com/
[16] http://php.net/
[17] https://httpd.apache.org/

Since the study used open source tools to develop the system and Stack Overflow developer community domain to identify digital heroes, it will be easy to customize and implement the system. The system can be also used in other domains like academic, medical and social communities.

5.5. Logical System Design and Specifications

The outputs of this stage are implementation-independent and concentrate on the requirements for the human computer interface. The main areas of activity are the definition of the user dialogues. These are the main interfaces with which the users will interact with the system. The logical system design specifies the main methods of interaction in terms of menu structures and command structures.

5.5.1. Input Design

As DHD platform is an automated system, it does not require any input from the user. Users only can search the specific digital hero or a group of heroes using a search form which has only three input fields. Visitors can search a digital hero using the name, hero category or country name.

5.5.2. Output Design

The outcome of the DHD web platform was designed after a careful analysis of the user experiences (UI/UX) in the web. All discovered digital heroes are listed and displayed with their name, Stack Overflow profile picture, button of contributions and hero category name. After navigating to the user profile, it will display details of contribution records that made this user a digital hero. There is a button to navigate Stack Overflow user profile as well.

5.5.3. Screen Layouts

DHD web platform was designed using open source web languages namely HTML, CSS, JavaScript and PHP. The study focused on user experience to design a better layout for the digital hero web platform. The platform analyzes data as a background process and displays output in the platform.

Homepage: DHD web platform homepage has a feature for searching digital heroes using hero category, name and country. The search form will redirect visitors to a new page where they will see the list of digital heroes based on their search filter. The page displays a maximum of 40 records at a time and it has Next and Previous buttons to see the rest of the records if it has more than 40 records. It also has navigation to browse other pages like About DHD, About Stack Overflow, About the study, Hero Criteria, DHD data processing model etc. as shown in Figure-6.

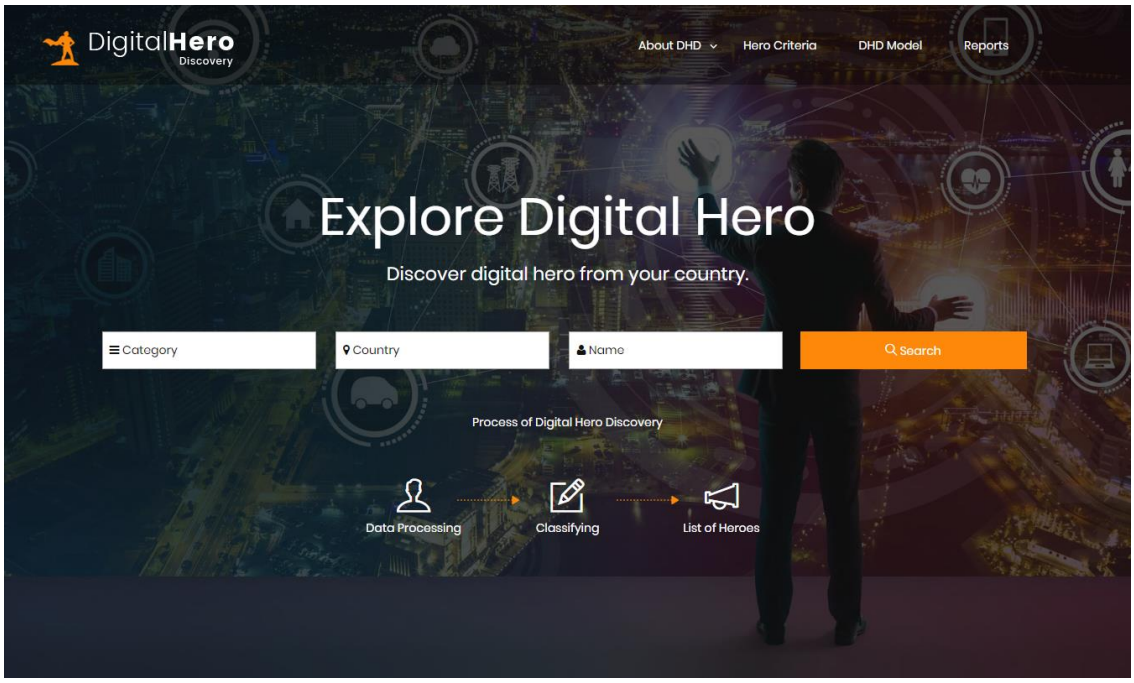

Figure 6: DHD Web Platform Homepage

Figure-7 shows the next section after the home page section. It displays the digital heroes as outcomes of the system. It shows the hero profile picture, name and country

that users used in Stack Overflow profile, hero category based on the contribution to the community, and a button to the detailed contribution page. This section initially displays 40 records and it has Next and Previous buttons to see other records as well.

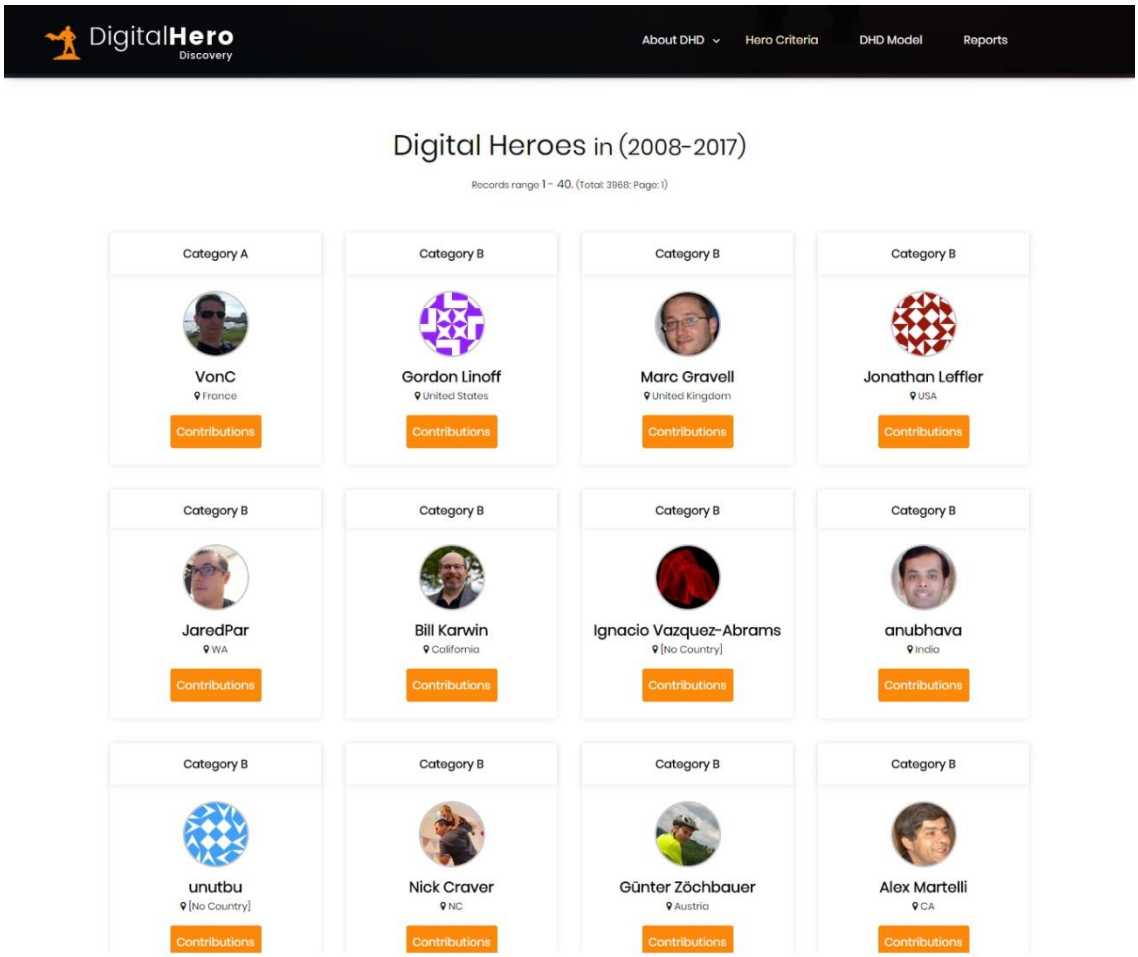

Figure 7: Discovered Digital Hero Display in the DHD Platform

Hero profile: When a visitor clicks on the ‘Contributions’ button from the home page, they will be redirected to the hero profile page. As shown in Figure-8, the hero profile page shows details of the records of the contributions that were provided by the digital hero in the Stack Overflow community forum. It shows the reputation number that was earned from Stack Overflow by contributing in the community, total gold, silver and bronze badges earned, recent 10 activities provided by the user and top 10 tags that the user contributed by time in the community.

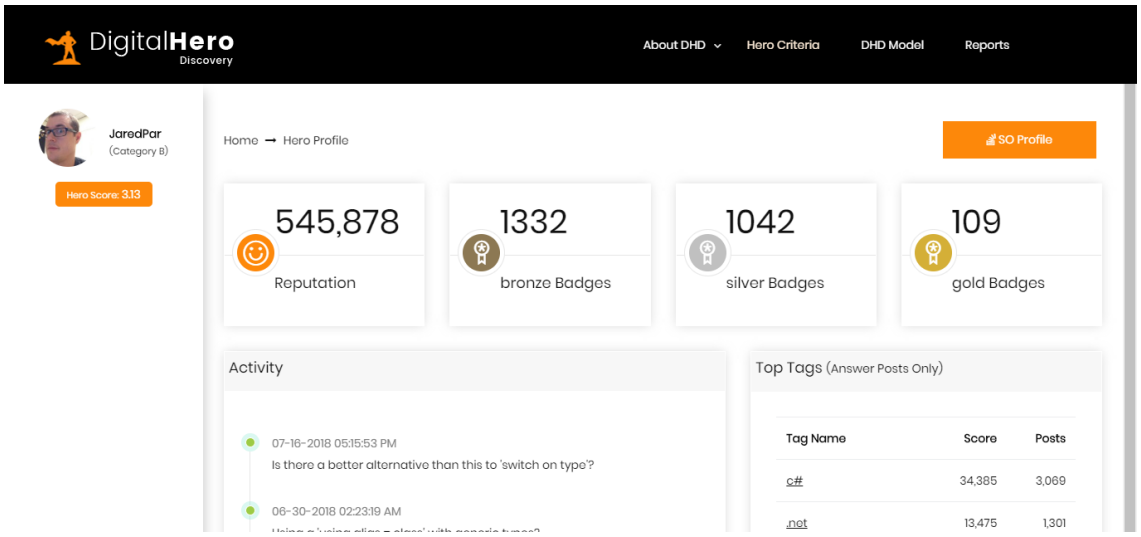

Figure 8: Digital Hero Profile Page of DHD Web Platform

Though DHD web platform analyzes data as a background process, it has a better user interface that enables users to know about the digital heroes that were discovered by analyzing Stack Overflow community forum users' information. The users get inspired by the digital hero's extra-ordinary contribution and sacrifice of their valuable time for the sake of helping community people.

5.5.4. Accessing DHD Platform

DHD web platform is an online web based automated system which can be accessed using any web browser with Internet connection. The URL to access DHD web platform is http://dhd.staritsoft.com/ (see Appendix 4).

5.6. Physical Design

The DHD data process model guided to design a physical database and set of conditions in order to analyze data and store the output in the database.

5.6.1. Database Schema and Structure

The structure of the DHD platform database is a relational data model. The tables relate to one another via unique primary keys of user id and foreign keys. Figure-9 shows a simplified database relational schema of the DHD platform database.

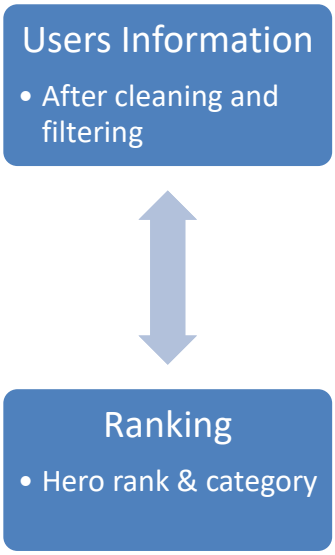

Figure 9: DHD Database Simplified Schema

5.6.2. Entities

Entities entail objects of interest to an application setup and which the particular application would be interested in keeping data about. DHD database only store user's information that was extracted from the dataset. Therefore, DHD system is made up of mainly 3 entities namely:

1. users_information
2. users_activities
3. hero_ranking

5.6.3. The Global Entity Relationship (ERD) Model

The entity relationship diagram (ERD) in Figure-10 shows the database design of the DHD web platform. To analyze users' digital activities the system used a dataset from Stack Overflow community forum. After cleaning the dataset, all relevant information was added into the DHD database for use by the next processes and displays the output of the system in the web interface. The database has 3 entities (see Section 5.6.2). All the 3 tables are related with each other based on the UserId.

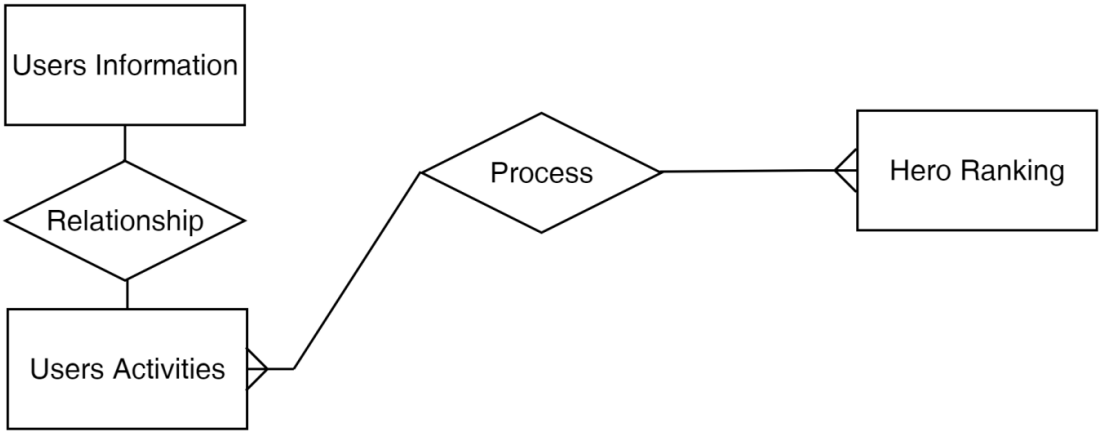

Figure 10: Global Entity Relationship (ERD) Model

5.6.4. Database Design and Data Columns of Tables

As shown in Figure-10 in the previous section, DHD platform uses 3 main tables to store digital heroes' information after analyzing the dataset. The following tables show the details about columns of the database tables.

Users Information Table: *users_information* table stores the information about a user that is relevant to the study which will be displayed on the user interface of the system. As shown in Table-13, this table mostly stores information related to the user profile including id, name, Stack Overflow profile link, profile creation date, user type, location etc.

Table 13: *users_information* Table

Field	Type	Null	Default
UserId	int(20)	No	
UserType	varchar(20)	No	
ProfileLink	varchar(100)	No	
CreationDate	datetime	No	
DisplayName	varchar(40)	Yes	
LastAccessDate	datetime	No	
Location	varchar(100)	Yes	Null
AboutMe	text	Yes	Null
ProfileImageUrl	varchar(200)	No	Null

Users Activities Table: The study created another relational table called *users_activities* to keep user activities related records that are relevant to scoring and classifying users using digital hero selection criteria. This table has a user id, user reputation, total number of answers, total up votes, down votes, total votes and number of years in the community.

Table 14: *users_activities* Table

Field	Type	Null	Default
UserId	int(20)	No	
Reputation	int(20)	No	0
AnswerCount	int(10)	Yes	0
UpVotes	int(10)	Yes	0
DownVotes	int(10)	Yes	0
TotalVotes	int(10)	Yes	0
Years_InThe_Community	decimal(3,1)	Yes	
QuestionCount	int(10)	Yes	0

Hero Ranking Table: This table keeps records related with digital hero classification and scoring digital hero based on their activities including total number of answers provided by the users in 3 years, monthly average answers within 3 years, last activity data, hero criteria score that the user earned including experience score, accuracy score, activity score, trust score and finally hero category.

Table 15: *hero_ranking* Table

Field	Type	Null	Default
Total_Ans_In_Three_Years	int(10)	Yes	0
Monthly_Avg_Ans	decimal(10,2)	Yes	0
LastActivityDate	Datetime	Yes	
ExperienceScore	int(5)	Yes	0
AccuracyScore	int(5)	Yes	0
ActivityScore	int(5)	Yes	0
TrustScore	decimal(6,2)	Yes	Null
HeroScore	decimal(6,2)	Yes	Null
HeroCategory	varchar(20)	Yes	Null

5.7. Systems Security

To ensure DHD web platform systems security, a number of security measures were put in place. The system was developed as an automated system. There is no user interaction with the core processes of the system. Everything is automated from collecting data from input, processing data based on DHD data processing model and providing output from the system. Users can only visit the platform to see and learn about the discovered digital heroes as well as their contributions. Users can also search for a digital hero based on hero's name, hero category and by country. The search form is secured by using form validation and user input validation approach.

5.8. Dataset Analysis and Discovering Digital Hero

After completing developing and testing the DHD web platform, the study used the dataset of Stack Overflow with 1,889,860 users' information, which constituted the sample size. The dataset is analyzed and processed in the background by the DHD web

platform. In order to analyze and discover digital heroes from the dataset based on digital hero selection criteria and proposed scoring procedure (see Chapter 4) using DHD data processing model that was developed by following the KDD process of data mining approach, the work inputted all the necessary information and it was processed as a background process of the DHD web platform. After data processing, the DHD web platform displayed a list of digital heroes in the platform as guided by the DHD data processing model (see Section 2.1).

5.8.1. Data Cleaning

In the first step of the data processing the platform cleans irrelevant data from the dataset and only relevant information is passed to the next step of the data processing model. Relevant information is data related with user's activities in the community forum such as user id, user name, Stack Overflow profile link, profile picture, location, total number of answers, total up votes, total down votes, creation date, last activity date and reputation score. In order to carry out data filtering, this step also considered some calculation of total votes (sum of up and down votes), number of years in the community based on creation date, accuracy rate (see Section 4.2.2), total answers of the last 3 years and average answers in each month (see Section 4.2.3).

5.8.2. Data Filtering

In this step, the study applied the digital hero selection criteria in order to filter qualified users as digital heroes. The criteria parameters applied for filtering were experience, trust, activeness, accuracy. For each criteria parameter, many users were eliminated from the qualified list as shown in Table-16.

Table 16: Qualified and Eliminated Heroes

Criteria	Qualified Users	Eliminated Users
Experience	396,854	1,493,006
Trusted	4,914	391,940
Activeness	4,276	638
Accuracy	3,231	1,045

Total records of user's information inputted in the system were 1,889,860.

Table-16 shows that after each filter of the selection criteria, lots of users were eliminated from the list of digital heroes.

Experience Filter: The study applied experience criteria as a first filter in the system and the output shows that 396,854 users were qualified and 1,493,006 users were disqualified. The qualified users in this criterion were sent to the next criteria filter.

Trusted Filter: The qualified users in experience criteria (396,854 users) were then passed through the trust filter (see Section 4.1.4) and only 4,914 users were qualified and 391,940 were eliminated.

Activeness Filter: In this step, 4,914 user's information was applied Activeness filter and as we can see in the Table-16, 4,276 users were qualified and 638 users were eliminated.

Accuracy Filter: As a last step of digital hero qualification, the 4,276 users were analyzed by Accuracy filter and 3,231 users were qualified and 1,045 users were eliminated.

After application of all selection criteria filters, only 3,231 users were qualified for all digital hero selection criteria from a total number of 1,889,860 users. As it is stated, all qualified digital heroes by criteria are not equal but their levels differ based on their different digital activities for the sake of helping community people (see Chapter 4).

After filtering, the study applied the proposed hero scoring procedure in the next step of data processing by the DHD web platform.

5.8.3. Data Scoring

In this stage, the study applied the proposed hero scoring procedure (see Section 4.2.5) to all users that qualified in all the selection criteria in order to differentiate them from one to another based on their contribution to the community. Table-17 shows the qualified users in each selection criteria within the score point range of 1 to 5.

Table 17: Users Who Qualified in Each Selection Criteria of Each Score Point

Score Point	Experience	Trusted	Activeness	Accuracy
5	976	3	1	1,109
4	1,041	10	0	888
3	687	15	0	675
2	493	97	5	331
1	34	3,106	3,225	228

Total qualified users were: 3,231.

5.8.4. Data Classification

In the data processing, classification was the last stage before the system produced the output. From the last stage, 3,231 users qualified and the study applied the proposed hero score procedure and proposed hero category in order to classify users in different categories (see Section 4.2.6). The output from the system is as shown in Table-18.

Table 18: Digital Heroes Based on Hero Score and Category

Hero Score	Qualified Users	Category
>= 4	1	A
3 – 3.99	45	B
2 – 2.99	2411	C
< 2	774	D

Total qualified users were: 3,231.

5.9. Conclusion on DHD Platform

DHD web platform is expected to discover digital heroes from Stack Overflow community forum user's contributions using data mining approach. The platform interface has been well developed, well designed in the sense of user experience (UX) and tested in terms of input, data processing and output and found to be accurate, hence it is ready for use. The platform is simple, open source, easy to use, capable of mining large number of user's information and it is compatible with any operating system as well as any device including desktop computers, tablets and mobile phones.

CHAPTER SIX - SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

6.0 Introduction

This chapter presents the summary of the findings, conclusions and recommendations of the study. Appropriate conclusions and recommendations were made on the basis of the research study findings and DHD web platform data analysis, design and development. Finally, suggestions for further research in the area under study were made.

6.1. Answering the Research Questions

The aim of the study was to analyze digital activities of Stack Overflow community user's activity in order to discover digital heroes based on hero selection criteria as well as developing a web platform that was guided by a data processing model using Knowledge Discovery in Databases (KDD) of data mining approach. The study had four objectives as outlined in chapter one guided by four research questions listed below:

1) What criteria set can be defined to identify digital heroes based on their digital activities?
2) How to develop a conceptual model that will guide the development of web platform?
3) How data mining approach can be used for developing web platform to discover digital heroes by mining data from their online activities?
4) What information is available in the developer community online forums and how to analyze this information to discover digital heroes?

A total of 45 respondents were sampled for the hero selection criteria establishment. The criteria were used to collect data related to the above research questions. A total of 1,889,860 registered users of Stack Overflow who contributed a minimum of one

answer were sampled for use by the data processing model. Using the hero selection criteria and collected user's information from the Stack Overflow, it was possible to answer all research questions.

6.2. Summary of Major Findings

This section summarizes the study findings based on the above research questions and data analysis in Chapter 4 and Chapter 5.

6.2.1. Establish a Set of Criteria to Define a Digital Hero Based on Their Digital Activities

This study established the digital hero selection criteria by summarizing expert's opinions which helped to develop the criteria. The study proposed four hero selection criteria parameters that were used to analyze Stack Overflow dataset in order to discover the digital heroes from the community forum members. Based on the expert's opinion, the four criteria parameters were minimum 6 years of experience in the community, 3 years of continuous activeness by contributing to the community activities, minimum of 80% accuracy in their contribution and minimum 20,000 reputation score to be considered as trusted by the community. The study measured trust by using Stack Overflow reputation score.

6.2.2. Develop a Conceptual Model of the Web Platform for Digital Hero Discovery

To analyze Stack Overflow dataset in order to discover digital heroes based on hero selection criteria, the study developed a conceptual model (see Section 2.1) that was guided by the KDD process of data mining approach. The model has 6 stages namely: data collection and input to the system that used a dataset from Stack Overflow; data

cleaning that works to clean unnecessary, noisy and irrelevant information for the study; data filtering stage responsible for filtering and calculating relevant data for the next stage using rules and conditions; data ranking stage for scoring a user based on their activities in the community and differentiating one user from another; data classification stage for categorizing digital heroes based on their score and proposed hero category; and finally the system provides the output of a list of digital heroes that is stored in the system database and displayed in the web platform. The model also guided the development of the DHD web platform.

6.2.3. Develop the Web Platform for Hero Discovery Using the Data Mining Approach

Guided by the conceptual framework, the study was able to develop a web platform using open source web development languages and design the relational database to store the relevant data. To develop the platform, the study focused on user interface and user experience (UI/UX) so that visitors to the platform can be able to get important information related to this study and view the discovered digital heroes as well as learn about their contribution to the digital community. The platform has a feature for searching a digital hero by country, hero name and hero category. It will give a visitor an opportunity to learn more about the digital heroes and get inspired by the contribution of these discovered heroes of the Stack Overflow community forum.

6.2.4. Collect and Analyze Information from Stack Overflow Developer Community Online Forum to Discover Digital Heroes Using the Web Platform

As mentioned earlier, this study used Stack Overflow community dataset. The dataset was analyzed to discover digital heroes based on a hero selection criteria and users'

contribution to the community. The study employed the KDD process of data mining approach to analyze the dataset (see Section 5.8) and discover the digital heroes. The dataset had 1,889,860 users information who contributed a minimum of one answer to the community and from that number of users, this study was able to discover 3,231 users as digital heroes who fulfilled all the hero selection criteria requirements.

6.3. Conclusion

Who does not like to be a hero? People always have intention of helping others. Just like in general life, our digital life with digital society affects us a lot. In this study, it is clear that discovering and recognizing someone as a digital hero would change our digital activity and inspire us with the intention of being heroes too. The method of inspiring community people of Stack Overflow to help other people is providing them with a reputation number based on their community activity and some badges named Gold, Silver & Bronze. So, we believe our study would inspire more digital users to increase good contribution for the community people.

The study established a set of hero selection criteria to measure people's digital contributions in the community with a view of defining someone as a digital hero. The established hero selection criteria are experience, accuracy, activeness and trust where a minimum of 6 years of experience in the community, 3 years of continuous activeness by contributing to the community activities, minimum of 80% accuracy in their contribution and minimum 20,000 reputation score to be considered as trusted by the community. A digital hero should have attained a certain level of community activity to be considered as a digital hero based on their digital contribution. The scope of this study was limited to only one community forum called Stack Overflow. However, four

proposed digital hero selection criteria can be implemented for other digital communities such as academic, medical and social communities.

It is obvious that the digital life is not exactly like the general life. In the digital community, there are millions of people from around the world communicating with each other using digital media. It is not easy to do human judgment on these millions of user's digital activities. Such digital activities require an automated system with rules and conditions in order to judge users digital activities and discover digital heroes. In this study, we used Knowledge Discovery in Databases (KDD) process of data mining approach to mine and identify digital heroes from millions of Stack Overflow users. We found data mining approach to be the best method to analyze huge volumes of data.

The study concludes that a web platform for digital hero discovery such as the one developed by this study, is a useful system for keeping digital heroes records for future generations. As we learn about general life heroes and get inspired by their extra-ordinary contribution to mankind, it is obvious that keeping historical records of digital heroes would inspire the future generation to do good deeds in digital life as well.

Finally, it is evident from the above findings that this study has achieved its aims and objectives. The findings indicate that by using the proposed criteria set and method of ranking to discover digital heroes would inspire people do to good activities in the digital community and inspire future generations as well.

6.4. Recommendations

The findings of this study demonstrate that discovering digital heroes using data mining approach is dependent on many components. The researcher therefore made the following recommendations to the forum's decision makers, researchers, data analysts,

community forums management teams and anybody interested in discovering digital heroes or even identifying members with the highest contributions emerging from the study findings.

This study was limited to Stack Overflow community forum and by using expert's opinion, the study was able to establish four hero selection criteria: experience, accuracy, activeness and trust in the community. It is recommended that these hero selection criteria could be enhanced by following the criteria development procedure of this study based on the domain of the data source. Enhancing criteria could give more accurate judgment of digital people activities and improve the discovery of digital heroes.

As mentioned earlier, this study used Stack Overflow community forum dataset to analyze and discover digital heroes based on their contribution in the Stack Overflow community forum. By following KDD process of data mining approach this study was able to develop a conceptual model of data processing. It is recommended that the model could have more stages of data processing based on data and the domain.

The developed DHD web platform used a dataset of 2008 to 2017 from Stack Overflow community forum. It contained information related to users of the community forum and their contributions for that period of time. It is recommended that the system could be customized to access and analyze real time data from the community forum and keep monitoring user activities and publish a list of digital heroes in real time.

This study was limited to discovery of digital heroes based on their general contribution in the Stack Overflow community forum. It is recommended that digital heroes be discovered and classified further based on their area of expertise. This way, it is possible

to differentiate one hero from another as well as appreciate their contribution in their specific areas of expertise. For example, in the Stack Overflow community there are different areas where experts are contributing including Java developers, PHP developers, .NET developers and C# developers. The DHD web platform could be customized to discover heroes in Java programing, PHP programing, .NET programing or even C# programming.

The developed DHD web platform currently works with only one domain of developer community forum called Stack Overflow. With a little modification, it is possible to use this digital hero discovery approach to discover heroes in other domains as well. There are digital community forums available where people are contributing and helping others based on their area of expertise. These include academic, medical and social community forums. It is therefore possible to use our approach to discover digital heroes and inspire more and more people in other community forums for the sake of increasing help to the digital community.

6.5. Suggestions for Further Research

This study of discovering digital heroes was conducted on a developer community forum called Stack Overflow. There is need therefore to carry out similar studies in other community forums and analyze their data types and structure in order to establish their hero selection criteria and discover digital heroes in those forums.

This study used the KDD process of data mining approach for data analysis. More research should be carried out with the view of discovering heroes using other approaches of data analysis and data process models.

REFERENCES

Alex. (2014). 14 Programming Communities for Developers, Hackers. Retrieved from https://codecondo.com/programming-communities/

Ary, D., Cheser, L., Sorensen, C., & Razavieh, A. (2010). *Introduction to research in education* (Eight). United State.

Bill Murphy Jr. (2014). 5 Qualities of Incredibly Heroic Leaders. *Inc.* Retrieved from https://www.inc.com/bill-murphy-jr/5-qualities-of-incredibly-heroic-leaders.html

Chauhan, D., & Jaiswal, V. (2016). An Efficient Data Mining Classification Approachfor Detecting Lung Cancer Disease. *2016 International Conference on Communication and Electronics Systems (ICCES)*. Retrieved from https://ieeexplore.ieee.org/abstract/document/7889872/

Cheung, Y. L., & Fu, A. W. C. (2004). Mining frequent itemsets without support threshold: With and without item constraints. *IEEE Transactions on Knowledge and Data Engineering*, *16*(9), 1052–1069. https://doi.org/10.1109/TKDE.2004.44

Chitraa, V. (2010). A Survey on Preprocessing Methods for Web Usage Data, *7*(3), 78–83.

Conway, P. (1996). Preservation in the Digital World. Retrieved from https://www.clir.org/pubs/reports/conway2/index.html/

Dietz, J. (2016). Online Communities vs. Forums vs. Portals vs. Knowledge Bases: What's the Difference? Retrieved from http://blog.socious.com/online-communities-vs.-forums-vs.-portals-vs.-knowledge-bases-whats-the-difference

Edward A. Fox. (1995). Digital libraries: introduction. *ACM SIGOIS Bulletin*, 8–10.

Eslake, S. (2006). THE IMPORTANCE OF ACCURATE , RELIABLE AND TIMELY DATA Discussion Paper prepared for a Group of ' Eminent Australians ' working with the Indigenous community of the Goulburn Valley , Victoria to assist in independently measuring and analysing the success o, (May).

Fayyad, U., Piatetsky-shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in, *17*(3), 37–54.

Fernandez, G. C. J. (2002). Discriminant Analysis , A Powerful Classification Technique in Data Mining Department of Applied Economics and Statistics / 204 University of Nevada - Reno Reno NV 89557.

Fraenkel, J. R., & Wallen, N. E. (2013). How to design and evaluate research in education. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Government of Bermuda. (2017). National Heroes Guidelines : Criteria and Selection Process, (441). Retrieved from http://communityandculture.bm/files/static_pages/1486655515Criteria and Selection Process (February 2017).pdf

Gray, B. (2004). Informal learning in an online community of practice. *Journal of Distance Education*, *19*(1), 20–35. https://doi.org/Article

Harper, F., & Raban, D. (2008). Predictors of Answer Quality in Online Q & A Sites. *Chi*, 865–874. https://doi.org/10.1145/1357054.1357191

Jayanthi, M. A., Kumar, R. L., Surendran, A., & Prathap, K. (2016). Research contemplate on educational data mining. *2016 IEEE International Conference on Advances in Computer Applications, ICACA 2016*, 110–114. https://doi.org/10.1109/ICACA.2016.7887933

Jurczyk, P., & Agichtein, E. (2007). Discovering authorities in question answer communities by using link analysis. *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management - CIKM '07*, (January 2007), 919. https://doi.org/10.1145/1321440.1321575

Kinsella, E. L., Ritchie, T. D., & Igou, E. R. (2016). Attributes and applications of heroes: A brief history of lay and academic perspectives. *Handbook of Heroism and Heroic Leadership*, 19–35. https://doi.org/10.4324/9781315690100

Laeeka Khan. (2017). 10 Reasons why Giving Back to Society is Important. Retrieved from https://www.lyceum.co.za/press-releases/10-reasons-why-giving-back-to-society-is-important

Liu, B. (1998). Integrating Classification and Association Rule Mining.

Maletic, J. I. (2000). Data Cleansing : Beyond Integrity Analysis 1, 1–10.

Mannila, H. (2000). Theoretical Frameworks for Data Mining, *1*(2), 30–32.

Margaret, R. (2008). SSADM (Structured Systems Analysis & Design Method). Retrieved from https://searchsoftwarequality.techtarget.com/definition/SSADM

Mark, S., Philip, L., & Tornhill, A. (2007). *Research Methods for Business Students*. *Pearson Education Limited 2*. https://doi.org/10.1007/s13398-014-0173-7.2

McNally, B. (2016). What Makes a Hero? Retrieved from http://www.huffingtonpost.com/barbara-mcnally/what-makes-a-hero_1_b_11836486.html

Miller, D. (2011). What makes a Hero? Retrieved from http://www.walb.com/story/14157521/special-report-what-makes-a-hero

Muhammad, D., Mohamudally, N., & Babajee, D. K. R. (2013). A Unified Theoretical Framework for Data Mining. *Procedia Computer Science*, *17*, 104–113. https://doi.org/10.1016/j.procs.2013.05.015

Philip Zimbardo. (2011). What Makes a Hero? Retrieved from https://greatergood.berkeley.edu/article/item/what_makes_a_hero

Pitta, D. A., & Fowler, D. (2005). Internet community forums: An untapped resource for consumer marketers. *Journal of Consumer Marketing*, *22*(5), 265–274. https://doi.org/10.1108/07363760510611699

R. J., F., & N. E., W. (2003). *How to Design and Evaluate Research in Education.* (5th Editio). New York.

Reader's Digest Magazine. (2011). What Is a Hero? Retrieved from http://www.readersdigest.ca/features/heart/what-is-hero/

Republic of Philippines. (2015). Selection and Proclamation of National Heroes and Laws Honoring Filipino Historical Figures. Retrieved from http://ncca.gov.ph/about-culture-and-arts/culture-profile/selection-and-proclamation-of-national-heroes-and-laws-honoring-filipino-historical-figures/

Republic of Rwanda. (2009). DEFINITION OF HERO. Retrieved from http://cheno.gov.rw/index.php?id=243

Saleemi, N. A. (2007). *Systems Theory and Management Information Systems Simplified* (2nd Editio). Nairobi: Saleemi Publications Ltd.

Select Business Solutions Inc. (2018). What is SSADM? Retrieved from http://www.selectbs.com/analysis-and-design/what-is-ssadm

Sensing, R. (2000). SEGMENTATION BASED ROBUST INTERPOLATION – A NEW APPROACH TO LASER DATA FILTERING.

Sharma, S. (2016). Data Preprocessing Algorithm for Web Structure Mining, 1–5.

Slegers, J. (2015). The decline of Stack Overflow. Retrieved from https://hackernoon.com/the-decline-of-stack-overflow-7cb69faa575d

Stack Overflow. (2018). About Stack Overflow website. Retrieved from https://stackoverflow.com/company

Stolfo, S. J. (1998). Real-world Data is Dirty : Data Cleansing and The Merge / Purge Problem Real-world Data is Dirty : Data Cleansing and The Merge / Purge Problem. https://doi.org/10.1023/A

Suba, S., & Christopher, T. (2016). An improved and efficient frequent pattern mining approach to discover frequent patterns among important attributes in large data set using IA-TJ-FGTT. *2016 IEEE International Conference on Advances in Computer Applications, ICACA 2016*, (1), 38–43. https://doi.org/10.1109/ICACA.2016.7887920

Tayyab Babar. (2014). 10 Traits of Sucessful Heroic Leaders. Retrieved from http://www.lifehack.org/articles/productivity/10-traits-sucessful-heroic-leaders.html

Techopedia.com. (n.d.). Structured Systems Analysis And Design Method (SSADM). Retrieved from https://www.techopedia.com/definition/3983/structured-systems-analysis-and-design-method-ssadm

Trueman, C. N. (2015). Structured Questionnaires. Retrieved from https://www.historylearningsite.co.uk/sociology/research-methods-in-sociology/structured-questionnaires/

Uzun, L. (2015). The Digital World and the Elements in Digital Communication and FL Learning. *Encyclopedia of Information Science and Technology, Third Edition*, (Lc), 2106–2113. https://doi.org/10.4018/978-1-4666-5888-2.ch203

APPENDICES

Appendix 1: Letter of Introduction

Dear Respondent,

I am a Masters student of Information Technology in Moi University. I've been doing a study about Discovering Digital Heroes from our Digital World.

The focus of this interview is to get expert opinion (By expert, researchers referring who know about the Stack Overflow community forum, understand its activities, and know about developer community people and their activities in the forum). The opinions will be used to finalize the criteria of digital hero so that it can guide to analyze digital activities and discover digital hero and keep their historical records in one place for future generation.

Study Title: Developing a Web Platform for the Discovery of Digital Heroes from Stack Overflow Developer Community Forum using Data Mining Approach.

Defining someone as a hero we need a proper criteria set to identify digital heroes. We start by defining criteria of selecting heroes. The defined criteria will be used to guide in collecting relevant information and analyze them to identify digital heroes based on their digital activities. This study will define someone as a digital hero in the Stack Overflow developer community forum by considering the following criteria: Experience, Accuracy, Activity and Trust.

Your assistance in that regard would be greatly appreciated. For any queries/clarifications, do not hesitate to contact me on e-mail: abdurrob.soyon@gmail.com

Yours faithfully,

Abdur Rob
Masters Student, Information Technology,
School of Information Sciences, Moi University

Appendix 2: Questionnaire for Experts to Collect Their Opinion on Hero Criteria

Experience - In any discipline, becoming an expert requires years of experience. In your opinion, what is the minim**u**m number of years required for a community forum user to be considered as a digital hero?

A. 1 to 5 years
B. 6 to 10 years
C. 11 to 15 years
D. 16 to 20 years
E. more than 20 years

Accuracy - How much accurate activity do you think a community forum user should have in an online community forum to be considered a digital hero?

A. 0 to 50%
B. 51 to 60%
C. 61 to 70%
D. 71 to 80%
E. More than 80%

Activeness - How long continuous period of activity do you think a digital hero should have worked in an online community forum?

A. 1 year
B. 1 to 2 years
C. 3 to 5 years
D. 6 to 10 years
E. 10 to 20 years

F. More than 20 years

Trust - How much reputation score do you think a member of Stack Overflow community should have to be declared a digital hero?

A. 10,000 to 20,000

B. 20,000 – 50,000

C. 50,000 – 100,000

D. 100,000 to 500,000

E. Over 500,000

Appendix 3: Sample PHP Source Code

Database Connection:

```
<?php
//DB details
$dbHost = 'localhost';
$dbUsername = 'root';
$dbPassword = '';
$dbName = 'db_dhd';

//Create connection and select DB
$db = new mysqli($dbHost, $dbUsername, $dbPassword, $dbName);

if ($db->connect_error) {
    die("Unable to connect database: " . $db->connect_error);
}
?>
```

Code for Import Data From Csv File:

```
<?php
//load the database configuration file
include 'dbConfig.php';

if(isset($_POST['importSubmit'])){
    //validate whether uploaded file is a csv file
    $csvMimes = array('text/x-comma-separated-values', 'text/comma-separated-values', 'application/octet-stream', 'application/vnd.ms-excel', 'application/x-csv', 'text/x-csv', 'text/csv', 'application/csv', 'application/excel', 'application/vnd.msexcel', 'text/plain');
```

```
if(!empty($_FILES['file']['name']) && in_array($_FILES['file']['type'],$csvMimes)){
    if(is_uploaded_file($_FILES['file']['tmp_name'])){
        //open uploaded csv file with read only mode
        $csvFile = fopen($_FILES['file']['tmp_name'], 'r');
        //skip first line
        fgetcsv($csvFile);
        //parse data from csv file line by line
        while(($line = fgetcsv($csvFile)) !== FALSE){
            //check whether user already exists in database
            $prevQuery = "SELECT Id FROM users WHERE Id = '".$line[0]."'";
            $prevResult = $db->query($prevQuery);
            if($prevResult->num_rows > 0){
                //update member data
                $db->query("UPDATE users SET Id = '".$line[0]."', Reputation = '".$line[1]."', CreationDate = '". date("Y-m-d H:i:s", strtotime($line[2])) ."', DisplayName = '".$line[3]."', LastAccessDate = '". date("Y-m-d H:i:s", strtotime($line[4])) ."', WebsiteUrl = '".$line[5]."', Location = '".$line[6]."', AboutMe = '".$line[7]."', Views = '".$line[8]."', UpVotes = '".$line[9]."', DownVotes = '".$line[10]."', ProfileImageUrl = '".$line[11]."', EmailHash = '".$line[12]."', Age = '".$line[13]."', AccountId = '".$line[14]."' WHERE Id = '".$line[0]."'");
            }else{
                //insert member data into database
                $db->query("INSERT INTO users (Id, Reputation, CreationDate, DisplayName, LastAccessDate, WebsiteUrl, Location, AboutMe, Views, UpVotes, DownVotes, ProfileImageUrl, EmailHash, Age, AccountId) VALUES ('".$line[0]."','".$line[1]."','". date("Y-m-d H:i:s", strtotime($line[2])) ."','".$line[3]."','". date("Y-m-d H:i:s", strtotime($line[4])) ."','".$line[5]."','".$line[6]."','".$line[7]."','".$line[8]."','".$line[9]."','".$line[10]."','".$line[11]."','".$line[12]."','".$line[13]."','".$line[14]."')");
            }
        }
        //close opened csv file
        fclose($csvFile);
        $qstring = '?status=succ';
```

```
        }else{
            $qstring = '?status=err';
        }
    }else{
        $qstring = '?status=invalid_file';
    }
}
?>
```

Hero List Display in the Platform:

```
<?php
include('header.php');
$sql = "SELECT UserId, ProfileLink, DisplayName, Location, ProfileImageUrl, HeroCategory FROM users WHERE AccLevel >= 80 ORDER BY HeroScore DESC LIMIT $MAX_DISPLAY_RECORDS;";
// echo $sql;
$query = mysqli_query($connect, $sql);
$rows = ( $query ) ? mysqli_num_rows($query) : 0;
?>
```

User Loop Section:

```
<?php
                if($rows > 0) {
                    while($record = mysqli_fetch_array($query)) {
?>
```

```
<div class="col-lg-3 col-md-6">
  <div class="widget author-widget">
    <div class="widget-title">
      <h5>
        <?php
        $categoryExplode = explode(' ', $record['HeroCategory']);
        echo '<a title="See more heroes of this Category" href="search-results.php?category='. trim(end($categoryExplode)) .'">'. $record['HeroCategory'] .'</a>';
        ?>
      </h5>
    </div>
    <div class="author-widget-body">
      <div class="thumb">
        <a href="user-profile.php?UserId=<?php echo $record['UserId']; ?>"><img style="width: 80px; height: 80px; border: 2px solid #c8c4c4;" src="<?php echo $record['ProfileImageUrl']; ?>" alt="img" class="img-responsive"></a>
      </div>
      <div class="info">
        <h4><a href="user-profile.php?UserId=<?php echo $record['UserId']; ?>"><?php echo $record['DisplayName']; ?></a></h4>
        <?php
        $location = explode(',', $record['Location']);
```

```
                    echo '<a href="search-results.php?country='. trim(end($location)) .'"><span><i class="fa fa-map-marker" aria-hidden="true"></i> '. ( trim((end($location))) ?: '[No Country]' ) . '</span></a>';

                    ?>

                </div>

                <div class="btn-group">

                    <!-- <a href="https://stackoverflow.com/users/6309/vonc" target="_black" class="success-btn" title="Stack Overflow Profile"><i class="fa fa-stack-overflow" aria-hidden="true"></i> SO Profile</a> -->

                    <a href="user-profile.php?UserId=<?php echo $record['UserId']; ?>" class="danger-btn">Contributions</a>

                </div>

            </div>

        </div>

    </div>

<?php } } ?>
```

Appendix 4: How to Install, Run and Access DHD Platform

DHD web platform is to analyze user's digital activities in the Stack Overflow developer community forum in order to discover digital heroes based on hero selection criteria and data process approach. The platform is developed using PHP and it is installed on the server-side, along with MySQL database and Apache web server software.

Installation

As the platform already hosted in a global hosting server, it does not require any installation to access it. But if anyone wants to install it in a local server or in a different hosted server, it will be required following steps:

a) To install DHD system, an Apache webserver, MySQL database and PHP are required and should be installed prior to installing DHD system. Make sure that Apache and MySQL are running.

b) Copy and paste the contents of *dhd* folder from the CD and paste it to the root of the web server either in Linux or Windows. The document root is /var/xampp/htdocs/ most Linux distributions and c:/xampp/htdocs/ in Windows. The root path could be very based on the operating system and server installation path.

c) Open the web browser and go to http://localhost/phpmyadmin/

d) Create a database called *dhd_db* and import the database that provided into the CD

e) After successfully import database, got ot http://localhost/dhd/ on the browser.

Accessing MUWEBCAMPUS From the Web

Currently, DHD platform is installed and accessible through the internet. It can be access by visiting the URL http://dhd.staritsoft.com/

Appendix 5: List of Publications

Authors	:	**Abdur Rob**, Nicholas Kiget, John K. Tarus
Paper Title	:	**Digital Hero Criteria Based on Digital Activities**
Journal Name	:	International Journal of Strategic Information Technology and Applications **(IJSITA)**
Publisher Name:		**IGI Global**
Status	:	**Submitted and Accepted**

Appendix 6: Research Budget

PROPOSED RESEARCH BUDGET		
Expenditure Description	**Cost per Item (USD)**	**Justification for Expenditures**
Domain for DHD Platform	30	A domain will be required to browse project online ($15/yr – 2 years)
Web Hosting	120	Hosing will be required to upload project to browse online ($5/mo – 24 months).
Teleconference line	250	A teleconference line will be required to conduct interviews with the project participants.
Office supplies	350	Paper, Photocopying, Binding, Printing etc.
Journal articles	300	Purchasing journal articles for literature review.
IBM SPSS Statistics	600	To analyze statistical research data it will be required to get SPSS Software for minimum of 6 months.
TOTAL	**1650**	**(USD)**