A Prototype for mapping of tweets on state services for decision support: a case of Huduma Kenya

Judy N. Ng'ang'ira
Faculty of Information Technology (FIT)
Strathmore University

A Prototype for Mapping of Tweets on State Services for Decision Support: A case of Huduma Kenya

Ng'ang'ira, Judy Nyakairu

A Thesis Submitted in partial fulfillment of the requirements for the award of a Degree in Master of Science in Information Technology (MSc.IT) at Strathmore University

Faculty of Information Technology

Strathmore University Nairobi

Kenya

June, 2018

Declaration

I declare that this research is my original work and it has not been previously submitted and approved for the award of a degree by any University. To the best of my knowledge and belief, this Thesis contains no material previously published or written by another person except where due reference is made in the research itself.

Name: Judy Nyakairu Ng'ang'ira

Registration Number: 083757

Signature ……...

Date ……..

Approval

The research of Judy Nyakairu Ng'ang'ira was reviewed and approved by the following:

Bernard Shibwabo, Ph.D.

Faculty of Information Technology,

Strathmore University

Joseph Orero, PhD

Dean, Faculty of Information Technology,

Strathmore University

Prof. Ruth Kiraka,

Dean, School of Graduate Studies,

Strathmore University

Abstract

The growing public participation in decision making regarding the management of State resources demands for a tool that support meaningful insight of the many aspects on environmental issues, for the development and evaluation of alternative management options. Twitter has become quite popular among researchers due to its massive volume in data thus drawing a great interest by the public service community to answer questions relating to the use and misuse of public service offices. However, despite the growing participation by the researchers using twitter as a public service misuse detection the State does not seem to optimize the opportunity that twitter offers for detecting and monitoring the services offered at Huduma Kenya, a one stop shop offering a variety of state services in almost all the counties. The main objective of this research was to demonstrate that twitter tweets can be dependably grouped based on state services selected keywords. The magnitude of state service tweets can be predicted with high accuracy. The method used includes various steps that can be summarized as first categorizing the groups on twitter and defining them. Second, finding out how each group pattern of activity contributes value in group participation. Thirdly, the identified users were invited to contribute in the interviews. Fourth, analysis of the interview results was carried out enabling the researcher to identify findings of ill-structured decisions in state services. Also, a mixture of related investigation similarity diagramming and grounded theory techniques were used to identify different benefit-related trends, patterns, and evolving relationships through all interviewees. Then, the data was sorted and compared by group type to discover which themes were most repetitively related per group. Moreover, to estimate on the generalizability of these results to the user population at large, access usage logs was required to determine usage levels.

Acknowledgements

Special gratitude is of thanking those who made this thesis a reality. I would like to thank my husband who is a blessing to me and for an overflowing love. I would also like to thank my supervisor. I would like to acknowledge my parents for molding my character. Most important to the Almighty God for without him I am nothing.

Abbreviations/ Acronyms

CEBs	-	Customer Engagement Behaviors
CoP	-	Community of Practice
EISA	-	Enterprise Information Security Architecture
GIST	-	Global Innovation through Science and Technology
ISD	-	Integrated Service Delivery
ITIL	-	Information Technology Infrastructure Library
MDAs	-	Ministries, Department and Agencies
OAUTH	-	Open Authentication
OECD	-	Organization for Economic Co-operation and Development
OSN	-	Online Social Networks
RSS	-	Really Simple Syndication
SABSA	-	Sherwood Applied Business Security Architecture
SNS	-	Social Networking Sites
SVM	-	Support Vector Machine
URL	-	Uniform Resource Locators

Definition of Terms

Customer Engagement Behaviors	"Customers' behavioral manifestation toward a brand or firm, beyond purchase, that results from motivational drivers." (Doorn et al., 2010)
Crowdsourcing	A sourcing model in which organizations use predominantly advanced internet technologies to harness the efforts of the virtual crowd to perform specific organizational tasks (Saxton, Oh, & Kishore, 2013).
Innovation	"The process of translating an idea or invention into a good or service that creates value or for which customers will pay" (business dictionary, n.d.)
Social Media	"Forms of electronic communication through which users create online communities to share information, ideas, personal messages and other content" (Merriam-Webster's online dictionary, n.d.)

Table of Contents

List of Tables

List of Figures

VT OMNES
VNVM SINT

Chapter 1 : Introduction

1.1: Background of the Study

As human beings we are interconnected through social groups, interlinked relationships and interactions on social media. One of the most potent ideas in the social sciences is the notion that individuals are embedded in thick webs of social relations and interactions (Borgatti et al. 2009). Individuals are constantly faced with information overload whereby one has a lot of information at their fingertips hence, many alternative choices to make a decision needed in performing daily tasks. A decision is a choice among many alternatives available to an individual (Sauter, 2011). It is therefore necessary to have some sort of mapping with the information found on social media such as Twitter, to help in decision support on State services.

Mapping is an operation that associates each element of a given set (the domain) with one or more elements of a second set (the range) (Oxford dictionary, n.d.). An event can be defined as something that happens at a given place and time, whereby the presence of participants, an environment such as twitter is available, and a time when information is shared will determine the existence of an event in this text. We can now build on this and say that twitter users are constantly having conversations, agreeing and disagreeing over facts and opinions, so it is interesting that this information is being tracked (McGonial, 2016). For a prompt and informed decision to be made, connecting the relationship of opinions in the tweet data and accelerate the time-to-value that is required for decision support by enterprises in today's world a tool for mapping is required.

The increased exponential growth of the internet usage especially social media applications such as Twitter provides a way in which the State and private companies can engage with the citizens. This form of interaction can be mapped and relevant opinions collected. Huduma Center Kenya is a program by the Government of Kenya that aims to transform public service delivery by providing citizen's access to various Public Services and information from one stop shop service centers. This is made possible through integrated technology platforms. Huduma Center Kenya is a state service that is also on social media site Twitter, where relevant opinion can be mapped

providing better insight into the way national government decisions are made to efficiently serve the Kenya citizens.

The use of mobile technology to access the web is ubiquitous around the world enabling interactive engagement of many activities that were previously restricted to the physical presence of all participants involved in an activity completion. Online communities are communicating at an instance on Twitter. Twitter content is known as tweets that users microblog which is a modern occurrence that refers to the dissemination of brief memos to some or all members of the source's social network over a particular web based facility (Kaplan & Haenlein, 2012) despite the geographical dispersion of a participant.

Tweets can be streamed from Twitter Application Programming Interface (API) but can be challenging for enterprises that would like to use this conversation data in making rational decisions on how the citizens will perceive the innovative services being provided. This is due to information overload during the interpretation of text thus requiring a way to filter this information. It is therefore beneficial to have some sort of intelligent tool that can support enterprises in making prompt and informed decision of the perception of the innovative service at hand. Through mapping text correlation or patterns of datasets, use of an algorithm to extract and enumerate the datasets, analyze and present the results in a more informative way that can be used to avoid ill-decision making on limited resources.

1.2: Problem Statement

For many years the Kenya government has struggled with the problem of poor service delivery in its public service (Ng'aru & Wafula, 2015). The accessibility to public services, facilities and amenities is an essential factor as a social indicator measurement. The government of Kenya has delegated most of the ministry service into a one-stop shop known as Huduma Kenya that are located country wide. Huduma Kenya also provides online services through its social media platform specifically Twitter. A lot of feedback is received from the citizens through tweets and it is challenging to collect, analyze and map this feedback when trying to make decisions. The use of an innovation management platform within Microsoft, Cisco's I-Prize competition was

conducted to engage the organization and the public to help generate and choose an idea for a new product (Bailey & Horvitz 2010; Tierney & Drury, 2013; Jouret, 2009).

Social media use is being rapidly adopted by business enterprises; by 2011 approximately 83% of Fortune 500 companies were using some form of social media to connect with consumers, which defines a Communities of Practice (CoPs) group as people with a common interest or practice (Wang, 2011; Naylor, Lamberton, & West, 2012; Muller, et al. 2012). Identifying interesting and useful contents from large text-streams is a crucial issue in social media because many users struggle with information overload (Yang & Rim, 2014).

1.3: Research Objectives

i) To investigate the factors relating to enterprise decision support.

ii) To analyze the current techniques, approaches and architectures used for mapping of aspects in social media

iii) To develop a prototype for mapping of tweets on state services for decision support.

iv) To test the functionality of the proposed prototype.

1.4: Research Questions

i) What factors relate to enterprise decision support?

ii) Which are the current technique, approaches and architecture used for mapping of aspects in social media?

iii) How can the mapping prototype be developed?

iv) How can the functionality of the proposed prototype be tested?

1.5: Justification

The sheer volume of user-generated content available on social networks allows for sophisticated environmental scanning (Cachia, Compañó, & Da Costa, 2007). In Africa, social media networking sites are becoming widely spread than it has ever been before and it appears that people's perception of this technology is diverse (Florunso, et.al 2010).

This research will therefore be important in attempting to address how classification of unstructured tweets can help to avoid ill-structured decision making. A survey of business use of social media tools found that in 30% of organizations, at least half of employees are actively engaged with enterprise social media technologies (The Chess Group, 2011).

1.6: Scope

Although they may be numerous tools engaging enterprises with social media, this research study is strictly built on Twitter. Tweets will be streamed and collected; the tweets are classified to specific datasets that are relevant to the study. The datasets are used to engage in quest to find meaningful patterns in information tweeted on State services. This can assist as an innovative way for the state services reputation advancing business opportunities and identifying risk on the type of service they are providing through the critical feedback they receive.The tweets collected are specifically focusing on Huduma Kenya services and how the Kenyan citizen perceives the state service. This aims at ensuring the training data fed into the intelligent agent is capable of recognizing Kenyan terms. English was the research language that was focused on.

Chapter 2 : Literature Review

2.1: Overview

During the investigation of the prerequisite for a prototype that maps state services tweets for decision support, immense achieve of data was audited under community detection on online social networks. The literature in this chapter investigates various algorithms and techniques including the challenges experienced and merits that can be combined to develop the proposed prototype. The use of social media networks for community detection is an active field in decision making analysis that has attracted much attention from organizations and individuals. Due to the massive amounts of data that flow through social networks and the relative ease of accessing this data, analysis of social networks has become a research topic of particular interest (Detrick & Hu, 2013). Sentiment analysis, sometimes called opinion mining, provides a means of automatically determining the attitudes or opinions of users via the content they have created (Falkowski, Barth & Spiliopoulou, 2008).

For many years the Kenya government has struggled with the problem of poor service delivery in its public service due to graft loopholes. In October 2013 the Jubilee Government embarked on the implementation of the Huduma Kenya Program as a flagship project in its Manifesto to address the requirements of the Kenyan Constitution on quality and access to Public Services. Huduma Kenya is a 'one stop shop' approach in reforming service delivery in the Kenya public sector (Ng'aru & Wafula, 2015; Oyugi, 2015). Huduma Kenya purposes to offer efficient government services at the accessibility of the public by combining integrated services inside one building, perhaps on the same floor, successfully making it promising for service pursuers to access it conveniently.

Online e-Huduma web portal is intended to deliver integrated services offered by various government ministries, departments and agencies (MDAs) and a combined and integrated channel Huduma payment gateway to facilitate ease of payment for government services, through postapay ((Ng'aru & Wafula, 2015). The introduction of Huduma Centers recognized substantial advancement in the delivery of government service through ICT. Instead of one hoping from one

office to another seeking for government services, he/she can get them from one office hence saving time and enhancing satisfaction (Psck, 2014).

Decision Support Systems (DSS) are computer-based systems that aid decision makers to study and resolve ill-structured decision problems by incorporating database management systems with analytical and operational research models, graphic and tabular reporting capabilities, and the professional knowledge of scientists, managers, and decision makers. A DSS includes three major components: A dialog subsystem, a database subsystem, and a model base subsystem (Watson & Sprague, 1993).

2.2: Social Media Empirical Data

Social media refers to the means of interactions among people in which they create, share, and exchange information and ideas in virtual communities and networks. Social media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content (Kaplan & Haenlein, 2010). Additionally, social media relies on mobile and web-based technologies to make highly interactive applications through which individuals and communities share, co-create, discuss, and modify user-generated content. It introduces substantial and pervasive changes to communication between organizations, communities, and individuals.

Twitter is a microblogging social media platform which allows its users to post "Tweets" that are up to 140 characters long. Social media refers to a set of online tools open for public membership that support idea sharing, creating and editing content, and building relationships through interaction and collaboration (Kim & Ko, 2012; Abrahams, Jiao, Wang & Fan, 2012). Many organizations are embracing the adoption of social media platforms such as Twitter into their business processes. The definition of social networking sites is via three necessary criteria that users must: "(1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system" (Boyd & Ellison, 2008).We use the term social media to refer to a group of Internet-based technologies that allows users to easily create,

edit, evaluate and or link to content or to other creators of content (Kaplan & Heinen 2010).The use of social media by businesses has seen an increase use lately. (Holtzblatt, Drury, Weiss, Damianos, Cuomo, 2013).

Figure 2.1: Social Media Landscape, showing the different Aspects of Social Media (Adapted from FredCavazza, 2016)

2.3: Application of Twitter in decision Support

Twitter is an ideal data source for decision support: its users, who number in the millions, publicly discuss events, emotions, and innumerable other topics; its content is authored and distributed in real time at no charge; and individual messages (also known as tweets) are often tagged with precise spatial and temporal coordinates. The use of twitter-specific linguistic analysis and statistical topic modeling to automatically identify discussion topics across a major city in Kenya can be incorporated into an innovative service prediction model. Twitter data improves innovative service prediction performance versus a standard approach based on kernel density estimation.

Twitter as a social media platform, has brought new capabilities such as micro blogging, Wikis, RSS feeds, social tagging, innovation challenges and electronic social networks (Azad, Faraj,

Kane, Majchrzak, 2013). One of the appealing phenomena of the microblogging service is the fact that certain occurrences of wide interest for a community of users produce a sudden increase of mentions in real-time as they unfold. Social media has four major potential strengths: collaboration, participation, empowerment, and time. Social media can be empowering to its users as it gives them a platform to speak (Bertot, Jaeger, Grimes, 2010). The affordability of Internet access has broken down the barriers of entry for anyone with the ability to have a mobile device that can access the internet to inexpensively get access to information that can be shared on twitter.

With this said then the creation of a mapping tool that was defined in Chapter one will be of great value to decision support key stakeholders that want to collaborate and share opinions with an insight on social media of what the citizens are saying on state services. The latest stage of ICT has made us so technologically dependent that the cell phones have become the necessity these days (Kalia, 2013). This shows that the affordability of internet enabled device especially mobile phones and tablets can give a citizens the same capabilities that a physical Huduma Kenya center partakes with technology and even enhance the communication capabilities of collaboration and participation of sharing opinions and enhance efficiency.

According to the McKinsey Global Institute report (2012) there are ten value "levels" in the application of social technologies adding value to an organization. The report states that eight of these levels are applicable to four stages of the value chain which includes product development, operation and distribution, marketing and sales, and customer service. The other two levels are applicable in the entire organization and the value chain.

Figure 2.2: The Ten value levels in the Application of Social Technology(Adapted from McKinsey Global Institute, 2012)

2.3.1: Social Media Effectiveness on Time

In terms of time, social media technologies allow users to immediately publish information in near-real time (Bertot, Jaeger, Grimes, 2010). According to Gartner Research (n.d.) Gartner Hype Cycle methodology gives you a view of how a technology or application will evolve over time. It may also be used to provide an insight on tweet opinions and access the behavior of citizens towards a particular state service.

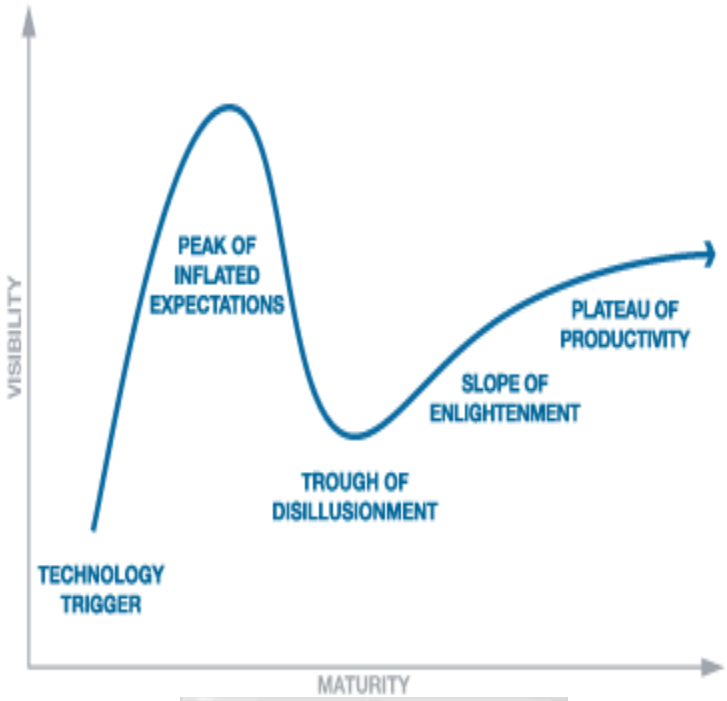


Figure 2.3: Gartner Hype Cycle (Adapted from Gartner, 2016)

The Gartner Research Group explains how the Hype Cycle five phases work:-

I. **Technology Trigger:** A possible technology advance boosts things off. Initial proof-of-concept stories and media interest trigger significant publicity. Frequently, no working products exist and profitable capability is not proven.

II. **Peak of Inflated Expectations:** Early marketing produces a number of success stories frequently complemented by tallies of failures. Some companies explore further others do not.

III. **Trough of Disillusionment:** Concentration decreases as trials and applications fail to deliver. Producers of the technology maneuver out or flop. Investments endure only if the surviving providers progress their products to the fulfillment of early adopters.

IV. **Slope of Enlightenment:** More illustrations of how the technology can profit the enterprise start to manifest and become more widely understood. Second- and third-generation products appear from technology providers. More enterprises sponsor pilots; conservative companies remain watchful.

V. **Plateau of Productivity:** Mainstream adoption starts to take off. Criteria for assessing provider viability are more clearly defined. The technology's broad market applicability and relevance are clearly paying off.

2.3.2: Social Media Foresight Methods for Open Innovation

According to Cachia, Compañó and Da Costa (2007), the applications of social media can be built on the relevance of three objectives of foresight methods, namely creativity, expertise and collective intelligence that can be formalized for open innovation:

I. ***Creativity*** is facilitated by collective action. Social media acts as a knowledge repository, which users with diverse knowledge use for novel combination and recombination during open ideation activities to facilitate ambidexterity.

II. ***Expertise*** is developed through the analysis of user-generated content and the ability to access local knowledge via engagement. Expertise generated from social media can be used to drive exploratory R&D as well as exploitative commercialization activities.

III. ***Collective intelligence*** emerges from the many-to-many interactions supported by social media during open innovation activities, which is applied to support ambidexterity, exploration, and exploitation across the sequential stages of the innovation funnel ideation, R&D, and commercialization

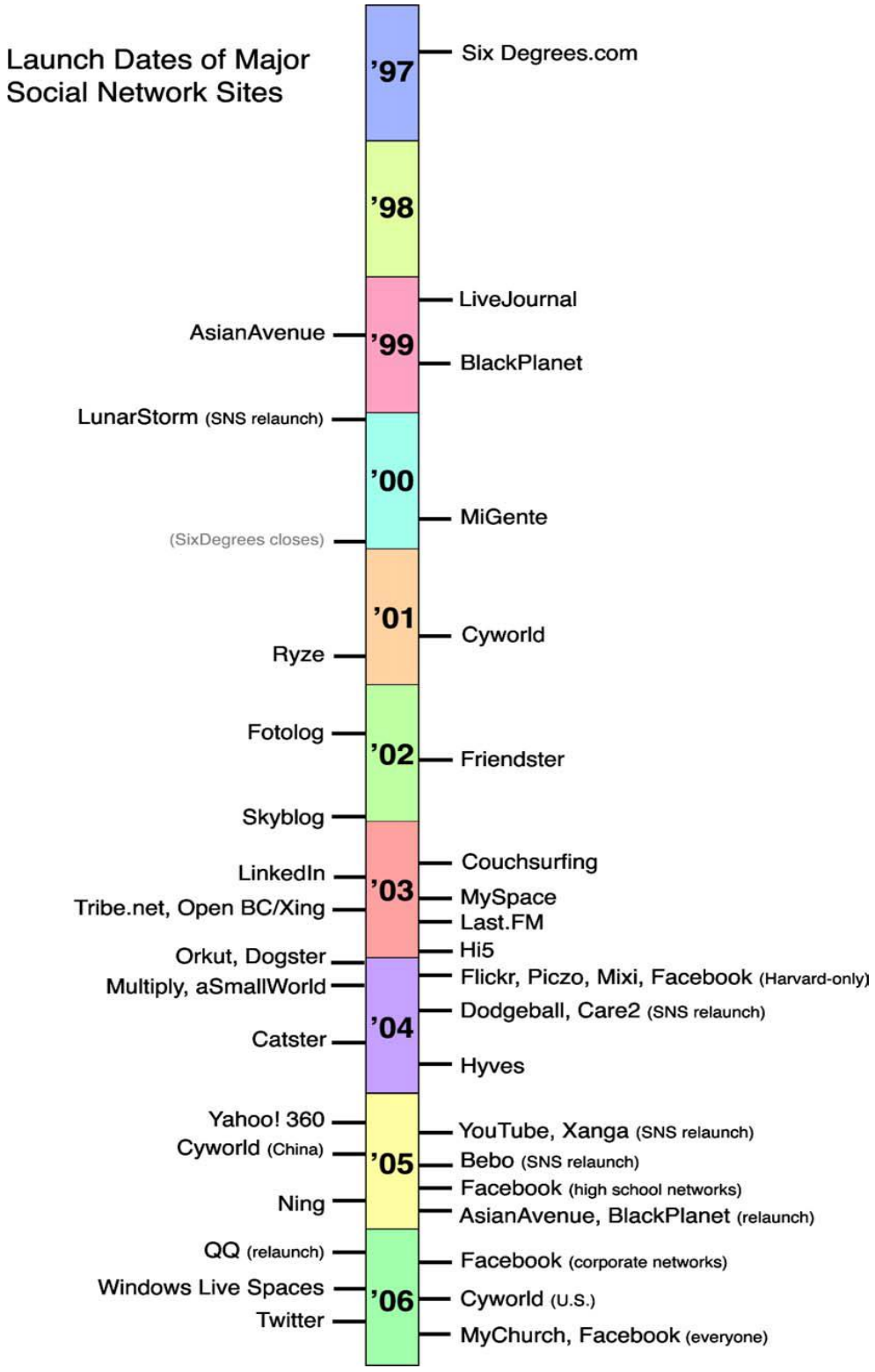


Figure 2.4: Timeline of the Launch Dates of Many Major SNSs and Dates when Community sites (Adapted from Boyd & Ellison, 2008).

2.4: Mapping Technology using Social Media

Networked communities when mapped, highlight the people and topics that drive conversations and group behavior, insights that add to what can be learned from surveys of communities. Maps of previously hidden networks on twitter contours the key people, groups, and topics being studied. Conversations on Twitter create networks with identifiable contours as people reply to and mention one another in their tweets. These conversational structures differ depending on the subject and the people driving the conversation. Six structures are regularly observed: divided, unified, fragmented, clustered, and inward and outward hub and spoke structures (Pew Research Group, n.d.).

Polarized Crowd network structure is only one of several different ways that masses and discussions can take form on Twitter. There are at most six unique structures of social media crowds which form subject on the topic being discussed, the material sources being quoted, the social networks of the people chatting about the topic, and the influencers of the discussion. Each has a different social structure and shape: divided, unified, fragmented, clustered, and inward and outward hub and spokes.

2.5: Application of Community Detection

One of the special interests in social network analysis is finding community structure. Community is a group of vertices that are tightly connected to each other and loosely connected with other nodes. Community detection is the process of network clustering into similar groups or clusters. Community detection has many applications including realization of the network structure, detecting communities of special interest, graph visualization (Masdarolomoor, Azmi, Aliakbary, & Riahi, 2011).

Community detection or clustering algorithms aim at finding groups of related nodes that are densely interconnected and have fewer connections with the rest of the network, these groups of nodes are called communities or clusters and they exist in a variety of different networks (Girvan & Newman, 2002).

2.5.1: Topic Detection in Collaborative Tagging Systems

Social Media applications contain a huge amount of hashtags that create the need for imposing organization with content tags that create spaces for collaboration. This can be openly attained by combination of tags that are built on the topics linked with the marketing strategy an organization wishes to get feedback on. Previous research work has attempted to derive meaningful clustering's of tags that match to topics of social importance. For instance, Begelman (2006) was amid the initial to apply community detection approaches, specifically spectral modularity maximization, to detect exciting tag clusters.

2.5.2: Tag Disambiguation

Tagging is unhindered on social media, hence many plentiful cases when a single tag in remoteness is not sufficient to convey the intended semantics. Therefore, tags need to be considered in context in order to disambiguate their meaning. Previous works have tried to address the issue of tag disambiguation by the use of community detection. Starting from a particular tag, Yeung, Gibbins, and Shadbolt (2009), derive several Social Media networks, e.g. a network of documents that have been tagged with the particular tag by the same user, and the community detection method of Newman is applied to extract communities of tags or documents (that eventually lead to tags) that correspond to the different senses of a tag. Using this technique resulted high performance compared to checking on some static eterior source of data such as WordNet. While collaborative tagging offers many merits over the use of controlled vocabularies, demerits such as the existence of polysemous tags.

2.5.3: User Profiling

Tailored search and references constitute an additional information recovery issue that can have advantage by usage of community detection. Precisely, clusters of tags have been proven to doing as current proxies of users' interests. Previous research by Gemmell, Shepitsen, Mobasher and Burke (2008) base the ranking performance of a personalized search mechanism on tag clusters outperforming conventional ranking schemes. The tag clusters were mined using a variation of

hierarchical agglomerative clustering. However, since this scheme requires manual parameter tuning that may have significant impact on performance (Gemmell, Shepitsen, Mobasher, & Burke 2008); a practical substitute would be to use some community detection to recognize tag clusters.

2.6: Sherwood Applied Business Security Architecture

In reference to a mapping carried out via the Internet, the need for security is paramount as a consideration when opinions among geographically dispersed citizens are concerned. According to a journal by Shariati, Bahmani and Shams (2010), a division of frameworks can be holistic or partial in Enterprise Information Security Architecture (EISA). They go ahead and mention most significant complete frameworks comprising Gartner, Sherwood Applied Business Security Architecture (SABSA), RISE frameworks. Shariati, Bahmani and Shams (2010), assert with Gartner having a theoretical essence defines three levels of concept which are conceptual, logical and implementation with three perspectives of business, information and technical for EISA framework. The disadvantage of Gartner's framework is that it does not offer a specific methodology for implementing EISA and has only given a general description of the structure and framework of EISA.

The separation of strategic planning initiatives, security solutions and activities that result in incompatibility of applying partial security solutions Functionality overlap between partial security solutions which led to the reduction of performance incomplete coverage of information security in the enterprise (Shariati, Bahmani & Shams, 2010). Matrix Model is an extension of the Information Technology Infrastructure Library (ITIL) and it is a six-layered architecture which entails of the horizontal layers of contextual, conceptual, logical, physical, part and the vertical layer of security service management. SABSA is further applicable and it's driven with a methodology, different from the Gartner framework. Its advantages are that it is generic and enterprises can use this as a starting point to defining the way forward in creating community detection on social media (Shariati, Bahmani & Shams, 2010).

SABSA	Assets (WHAT)	Motivation (WHY)	Process (HOW)	People (WHO)	Location (WHERE)	Time (WHEN)
Contextual	The Business	Business Risk Model	Business Process Model	Business Organisation and Relationships	Business Geography	Business Time Dependencies
Conceptual	Business Attributes Profile	Control Objectives	Security Strategies and Architectural Layering	Security Entity Model and Trust Framework	Security Domain Model	Security-Related Lifetime and Deadlines
Logical	Business Information Model	Security Policies	Security Services	Entity Schema and Privilege Profiles	Security Domain Definitions and Associations	Security Processing Cycle
Physical	Business Data Model	Security Rules, Practices and Procedures	Security Mechanisms	Users, Applications and User Interface	Platform and Network Infrastructure	Control Structure Execution
Component	Detailed Data Structures	Security Standards	Security Products and Tools	Identities, Functions, Actions and ACLs	Processes, Nodes, Addresses and Protocols	Security Step Timing and Sequencing
Operational	Assurance of Operational Continuity	Operational Risk Management	Security Service Management and Support	Application and User Management and Support	Security of Sites and Platforms	Security Operations Schedule

Figure 2.5: The SABSA Matrix (Adapted from SABSA Organization, 2015)

2.7: Association for Information and Image Management's (AIIM) model of Collaboration

The collaboration of human is complex yet can be very rewarding in a working environment. To address on the way to enhance the required effort, resources and support AIIM lifecycle model is an eight stage recursive loop.

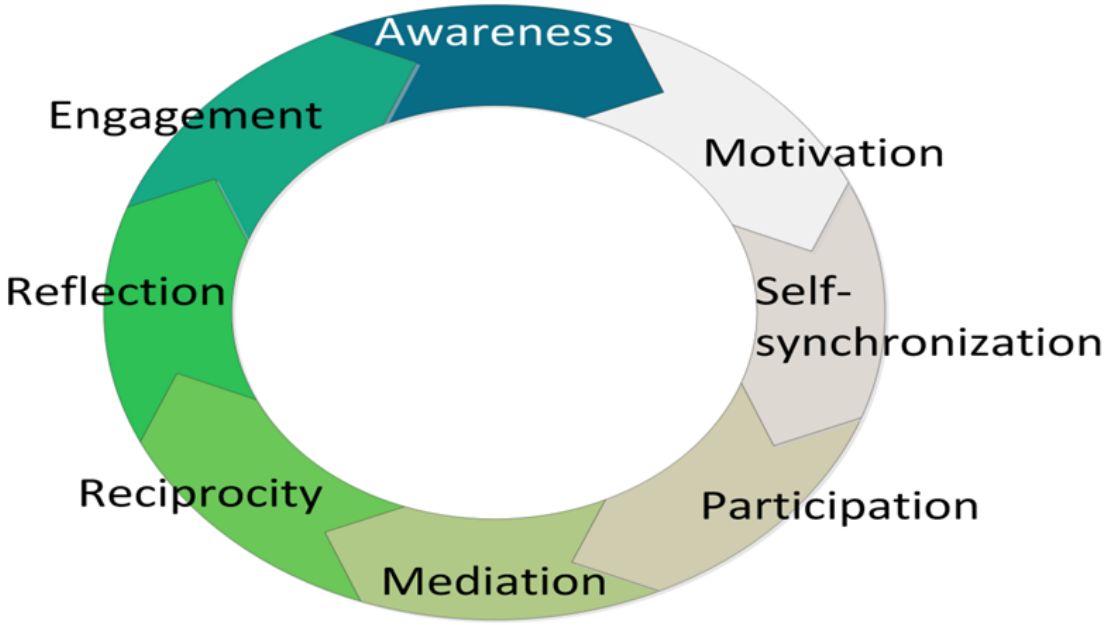


Figure 2.6: The Association for Information and Image Management's (AIIM) (Adapted from AIIM's Life Cycle Collaboration Model, 2013)

Awareness is a point where entities are formed by people with a shared drive. This makes a good beginner point for the sequential movement during collaboration though motivation being an initiative to gain agreement in problem development or solution raises question if it should be the next point or the end point. The cycle suggests that engagement is the end point, yet when collaboration begins to have engagement more ideas arise, making it a questionable ending point. Mediation the negotiation occurs finding a middle point. Reciprocity the sharing of findings from outcome happens in return through reciprocity Reflection critical thinking and the consideration of the available alternatives. Engagement occurs proactively than reactively.

2.8: Three P's and a G over T Collaboration Framework

This framework is for evaluating and managing organizational. This is the "Three P's and a G over T Collaboration Model."

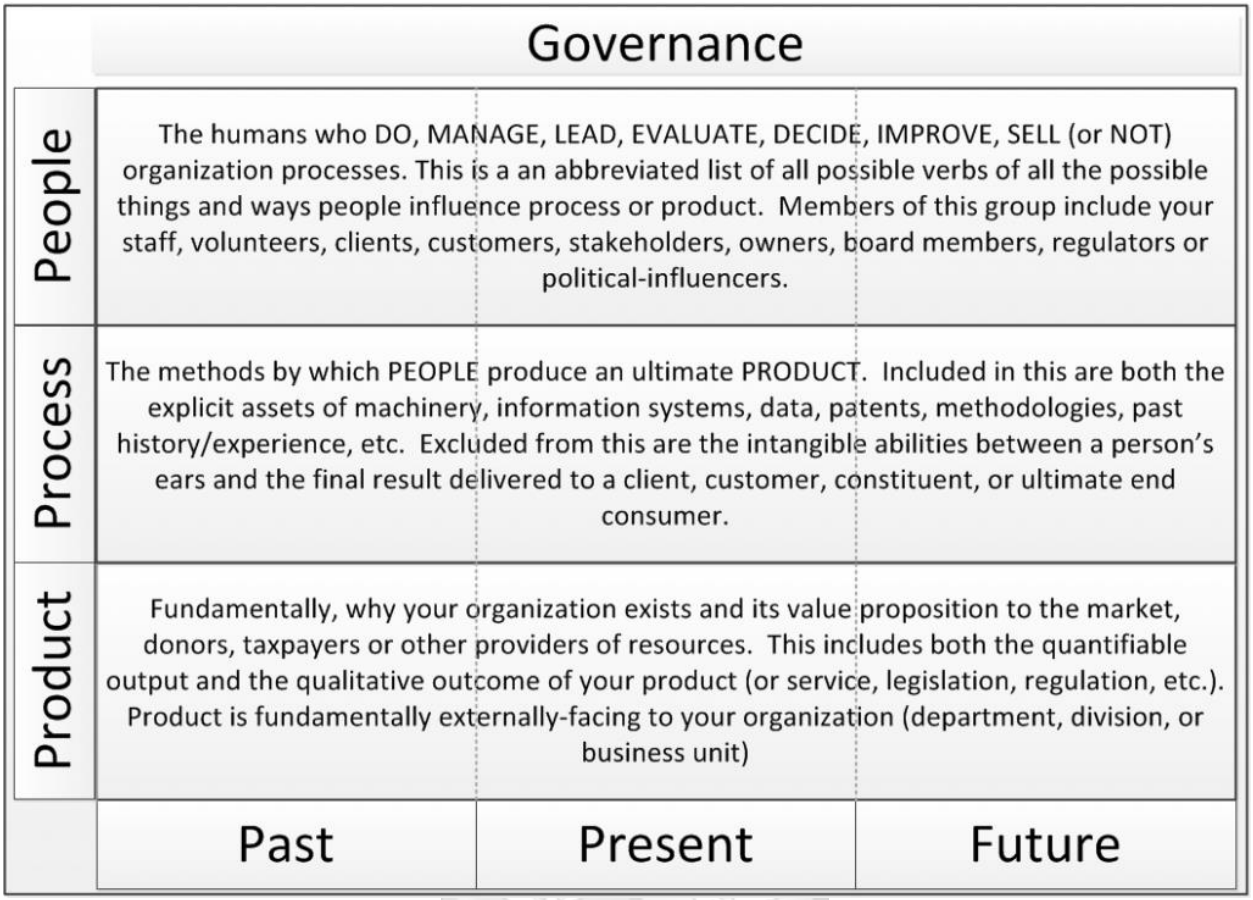


Figure 2.7: Three P's and a G over T a Collaboration Framework (Adapted from Collaboration is it Hard Wired, 2013)

On the x axis the model considers the time used during quality management when discussing and the y axis the model consider the people, process and product that are the fundamentals of an organization foundation.

2.9: Conceptual Framework

Independent variables are Huduma Center one stop shop and the dependent variables are Reliability, scalability, Heterogeneity Evolution and evaluation. Such a framework helps to ensure the reliability of public services by informing citizens of their rights and by giving them networks of amends and quality assurance. Statements of citizens' rights might also promulgate basic service and process standards, e.g. the Citizens' Charter that existed in the United Kingdom. The

assessment of citizens' rights recognition has become even more compelling in a context of decreasing trust in national governments and in leadership. Few countries have a common definition of patient or taxpayer rights, let alone a standardized regulatory framework for the implementation of complaint practices (OECD, 2013).

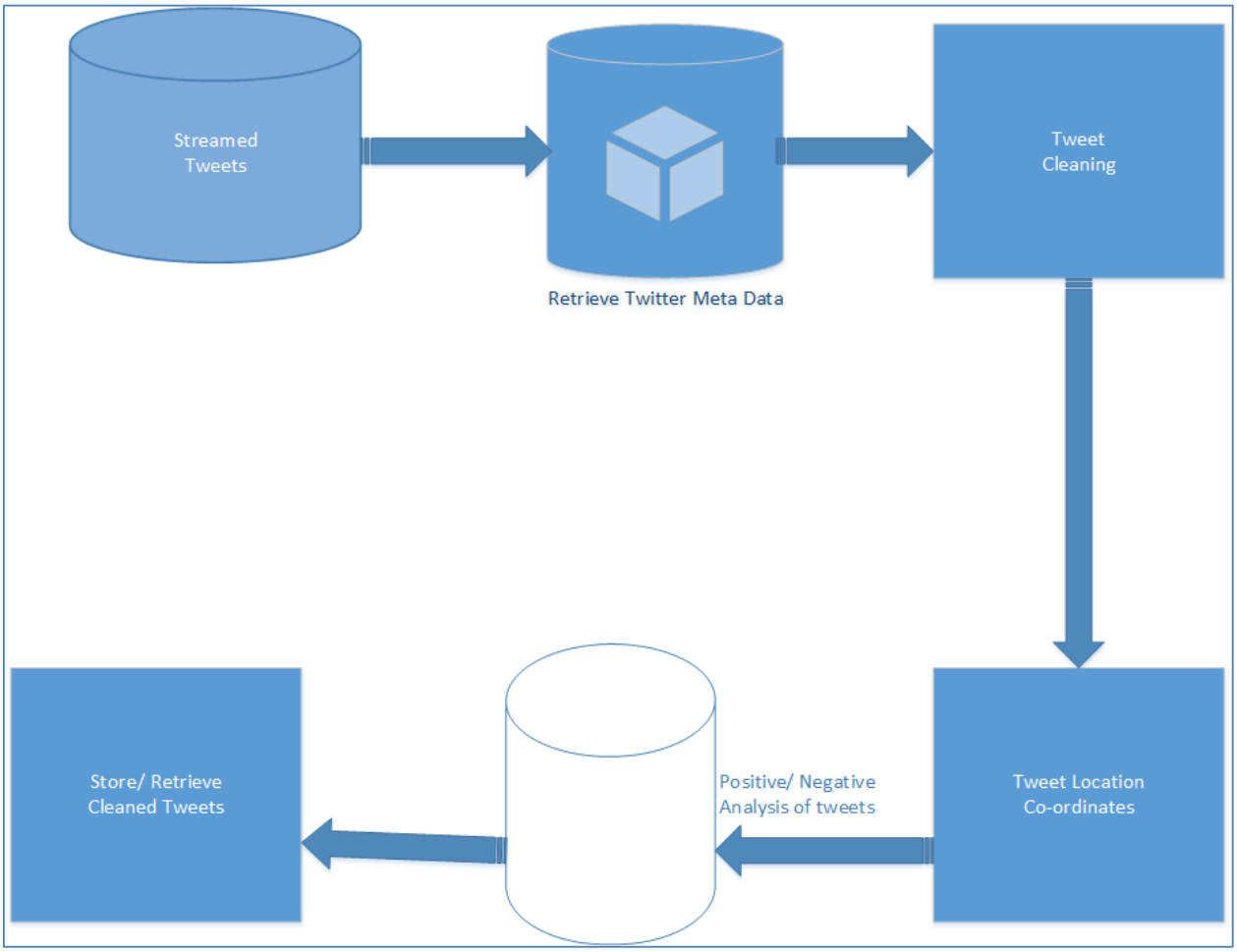


Figure 2.8: Proposed Conceptual Model

Chapter 3 : Research Methodology

3.1: Introduction

In this chapter, the research design, process and methodology used for this research are discussed. The population selected was on Huduma Kenya tweets. To address the research questions, Twitter data was first collected. This data was cleaned up by applying various training datasets. This was performed using community detection, and a prototype was designed to leverage on mapped state service tweets for decision support with relevant opinions offered on Huduma Kenya.

3.2: Research Design

The research aimed at developing a prototype for mapping Huduma Kenya state services for support in decision making. This research is applied research. Applied research has a practical problem-solving emphasis (Cooper & Schindler, 2014). Both qualitative and quantitative research method have been used in this research.

In this research through the use of a polarity check of negative and positive tweets on Huduma Kenya services, the need to rectify the system that is resulting in lost sales from Huduma Kenya users or an opportunity to increase stakeholders wealth through acquiring creative ideas from positive tweets on Huduma Kenya.

"Quantitative information go through a process that is primarily aimed at transforming the information into numerical values, called codes, so that the information can be easily analyzed, either manually or by computers" (Kumar, 2011). The use of qualitative research was used in this research.

3.3: Model development Population

To develop the proposed conceptual model, the following steps were followed:

I. Data was streamed for collection
II. Data was cleaned , categorized and analyzed

III. Development of the prototype

IV. Test of the prototype

3.3.1: Obtaining Data

This study's data was gathered through streaming data from the Huduma Kenya twitter streaming API. Information was also gathered from matching datasets with trained word list stored in a CSV file.

3.3.2: Data Pre-processing

The data collected contained gaps and duplicates thus it required to be cleaned through pre-processing. Cleaned data was processed through SQLite3 database and analyzed in accordance with the outline laid down for the purpose of deriving meaning from the trained dataset.

3.3.3: Development of the Prototype

Twitter API Libraries provided a connection to extract tweets. Python 2.7, an open source software was used that provides an environment for the development of the prototype. The use of Plotly library, to print the graphs was also used it is an open source python library. SQLite3 Libraries provide the database management system for the prototype.

3.3.4: Testing of the Prototype

For the measuring instrument to be valid the measuring instrument must be recognized to precisely measure a situation as intended by the researcher (Cooper & Schindler, 2014). The prototype testing was undertaken by use of training dataset. The training data was set from the set of data obtained online. The level of efficiency obtained by the prototype was assessed from the training dataset and the polarity data.

3.4: Population

Cooper and Schindler (2014) describe a population as the total collection of elements whereby references have to be made. The research target population was Tweets from all the 41 Huduma Kenya.

In general, strategies for detecting Huduma Kenya services posts using Twitter can be grouped into four natural language processing (NLP) methods:

i) Supervised grouping of words that relate and represent Huduma Kenya.
ii) Unsupervised models for one-stop shops using service techniques such as state service Topic Aspect Model
iii) Specific word counting, tracing geographic tweet dissemination
iv) Relating tweet contents with the Twitter location information.

In this research, we combine the above mentioned methods with the implementation of a Twitter streaming search tool designed to collect data and map its location.

Each tweet in the employed dataset was classified into either neutral, negative or positive polarities.

3.5 Filtering of Tweets

The research considers two main categories of tweets both taken from the filtered tweets: these categories were as follows:

i) Category 1: Huduma Kenya employees
ii) Category 2: Huduma Kenya clients

The study's data was gathered through streaming data from the twitter streaming API. Information was also gathered from matching datasets with trained word list stored in a CSV file.

3.6: Research Quality

The competence of the research design and the quality of tools and techniques used to attain and examine the data are dire measures since the quality of research project is dependent on methods. A measuring instrument that was reliable was required for data collection and to deliver dependable results Python software was used in collection of data, and thus the data was to be useful for the research.

3.7: Ethical Issues

The researcher addressed the participants on the purpose of the study when administering the data collection tools and also explained that with permission all information collected was treated as confidential and not infringing on their privacy rights by participating in the research. Also the researcher did not force participation to the study for the purpose of enough responses.

Chapter 4 System Design and Architecture

4.1: Introduction

In this section, we will outline the design architecture of the developed prototype of mapping Huduma Kenya tweets for decision support. This architecture will follow from the conceptual model shown in Figure 2.8. This section shall cover the interaction between the users and the developed prototype, the components of the developed prototype and the interaction of the various components of the developed prototype. This was modelled by use of Use case diagrams and sequence diagrams.

4.2: Requirement Analysis

Based on the objectives as well as the user requirements this section outlines the various requirements that will be met through this research.

4.2.1: The Functional Requirements

The functional requirements are as follows:-

I. Connect to Twitter
II. Collect Tweets related to Huduma Center Kenya based on specific keywords
III. Ensure the collected tweets have no duplicates.
IV. Retrieve the meta data of each tweet along with the content and location
V. The information stored in the database has the following information: Username, Geo-location, tweet content, search key word.

The need for a database to store the collected tweets was as follows:-

I. Created an SQLite3 database to store tweets related to Huduma Center and the co-ordinates
II. Performed tweet cleaning and stored in a database table.
III. Performed analysis and classification of tweets and compared to a polarity score of positive or negative
IV. Linked the tweets to a map geo-coding that shows visual maps physically.

4.2.2: The Non-Functional Requirements

The non-functional requirements are as follows:-

I. Registering the application prototype on Twitter to have access keys.
II. Provide a consumer key, consumer secret and access token to Twitter Streaming API for accessing data.
III. SQLite3 database configuration.

4.2.3: Usability Requirements

Since the use of this prototype requires access to Twitter, the user needs to have access to the internet and should at least have a basic understanding of how Twitter works in order to stream huduma Kenya tweets.

4.2.4: Supportability Requirements

The prototype should run on any smart device that has the required software installed.

4.2.5: Software Requirements

I. Plotly Libraries used to print graphs
II. Python 2.7 an open source software that provides and environment for the development of the prototype

III. SQLite3 Libraries provide the database management system for the prototype

IV. Twitter API Libraries providing a connection to extract tweets

V. Internet connectivity to stream tweets

4.3: System Architecture

Figure 4.1 presents the system architecture.

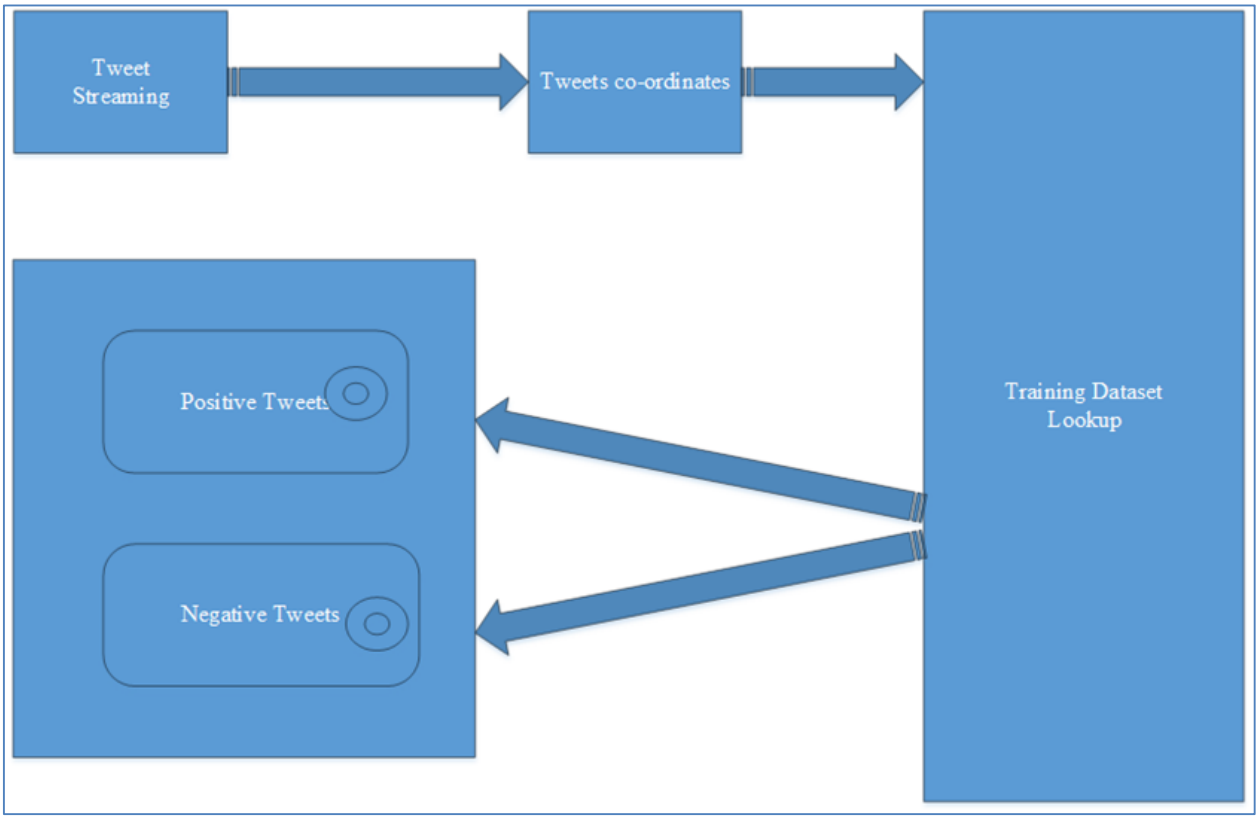


Figure 4.1: System Architecture

4.4: Use Case Diagram

Figure 4.2 describes the user's interaction with the prototype. The use case diagram comprise of the actor this is something with a behavior or a role. Hence in this research the actor was the prototype user. The boundary represents the limits to which the system operates.

Use Case Description

Use Case: Obtaining Huduma Kenya Tweets for streaming

Primary Actor

System User

Pre-condition

Internet access is available

Post Condition

The tweets streamed are comments on huduma Kenya services

Main Success Scenario

Actor intention	**System Responsibility**
1. User logs onto Twitter	
2. Retrieves huduma Kenya Tweets	
	3. Prints Tweets Co-ordinates
	4. Connects to database
	5. Stores/ retrieves cleaned Tweets
6. Views displayed graphs	
7. Exits Twitter	

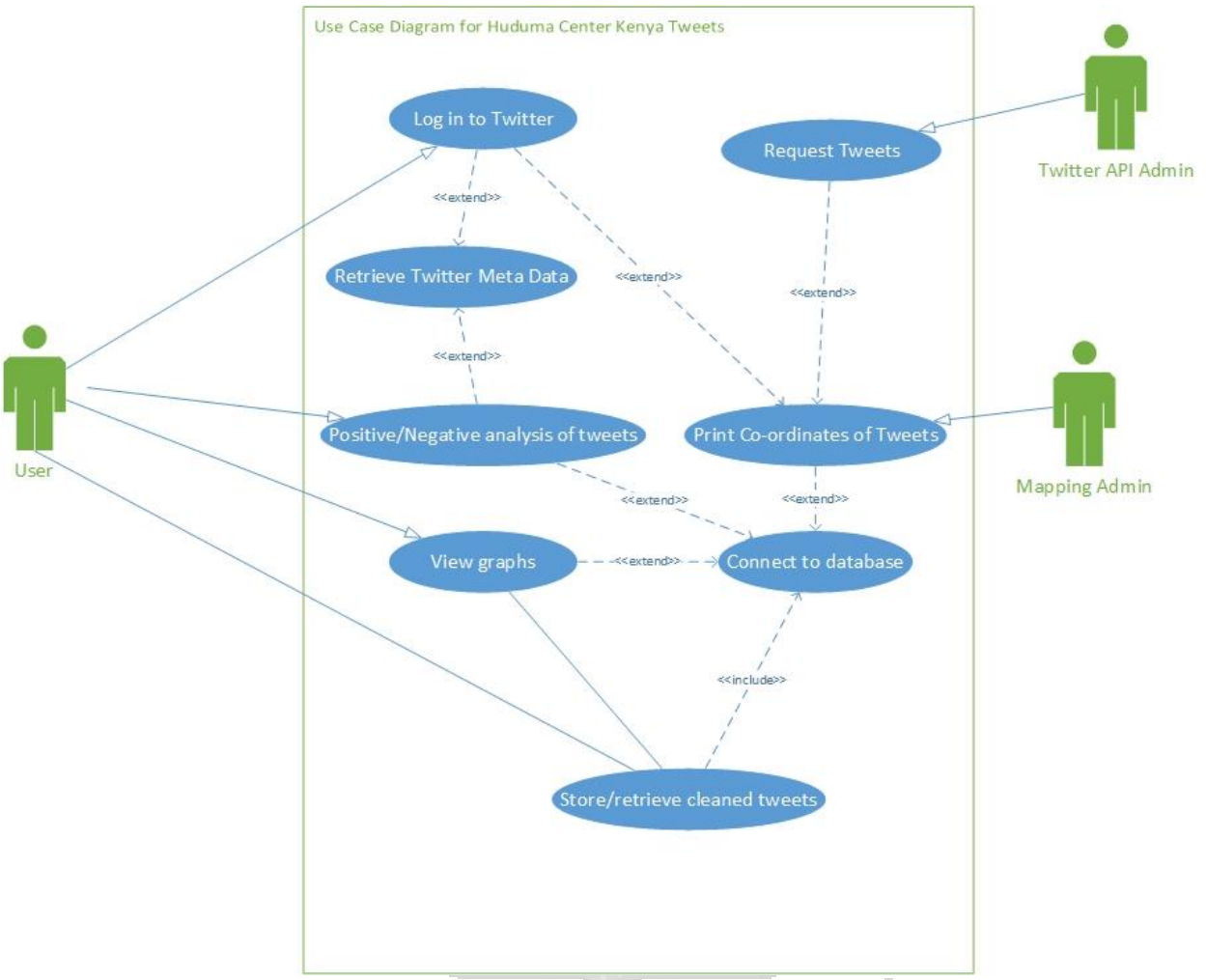


Figure 4.2: Use Case Diagram

4.5: Sequence Diagram

A sequence diagram is part of an explanation on how a system behaves. System Sequence Diagrams (SSD) are were used to show pictures of a scenario use case on the respective interactions.

4.5.1 Sequence Diagram for Data Collection

This is the initialization stage in the prototype. The user activates the use through running the python code. This is defined by Figure 4.3.

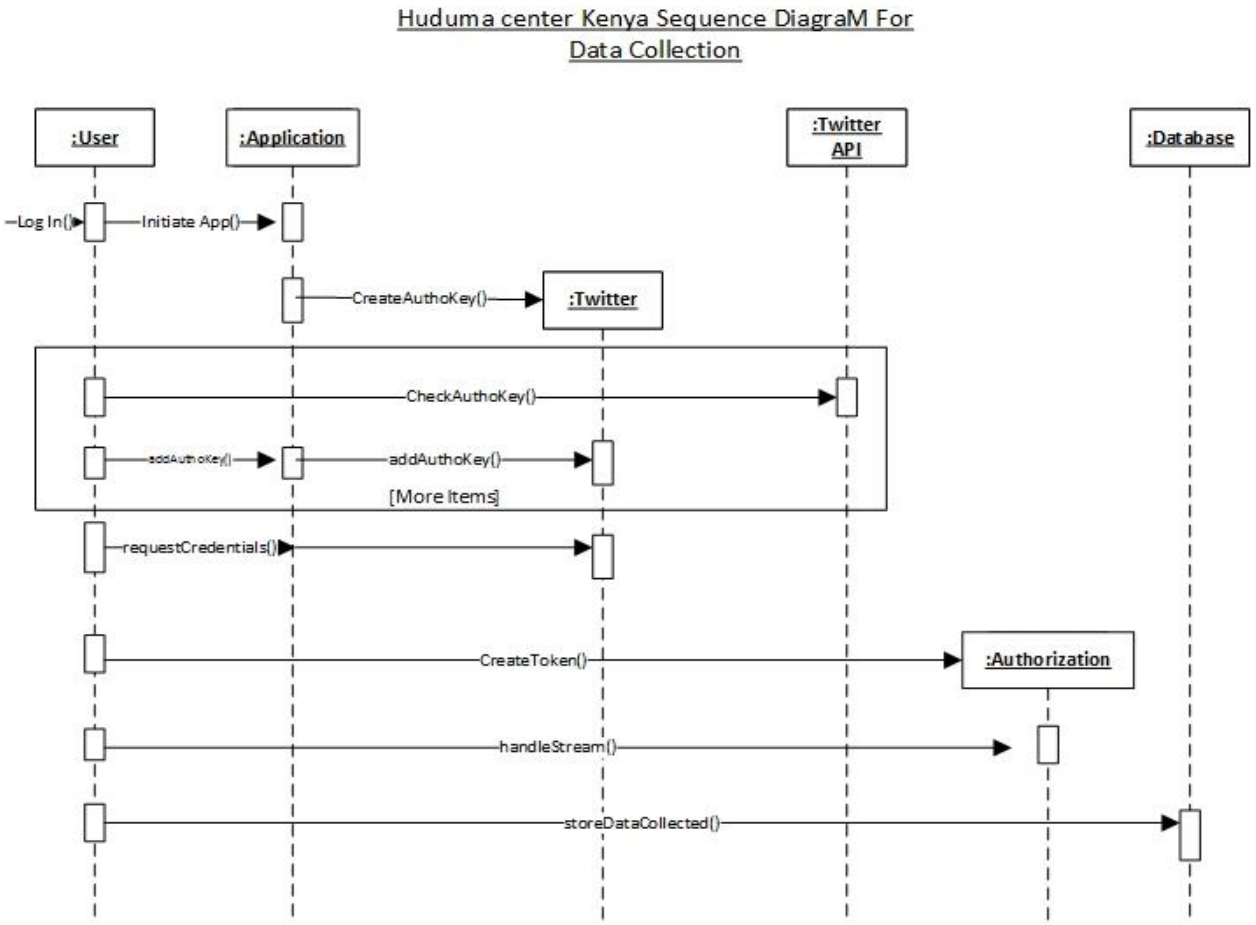


Figure 4.3: Sequence Diagram for Data Collection

4.5.2: Sequence of events

I. The user activates the process

II. The connection to the database is established by the prototype

III. The prototype establishes a connection to the Twitter Streaming API

IV. The prototype establishes credentials to access the API

V. The streaming of data is established by the API

VI. The data is written into the database

VII. The user terminates the system

4.5.3: Sequence Diagram for Cleaning Tweets

For the twitter data to be cleaned, the author developed datasets that were used from a data file to remove unwanted content. This is useful in classifying raw data correctly. Therefore, the data need to be totally accurate before the data cleaning takes place. Data cleaning sequence diagram is shown in Figure 4.4.

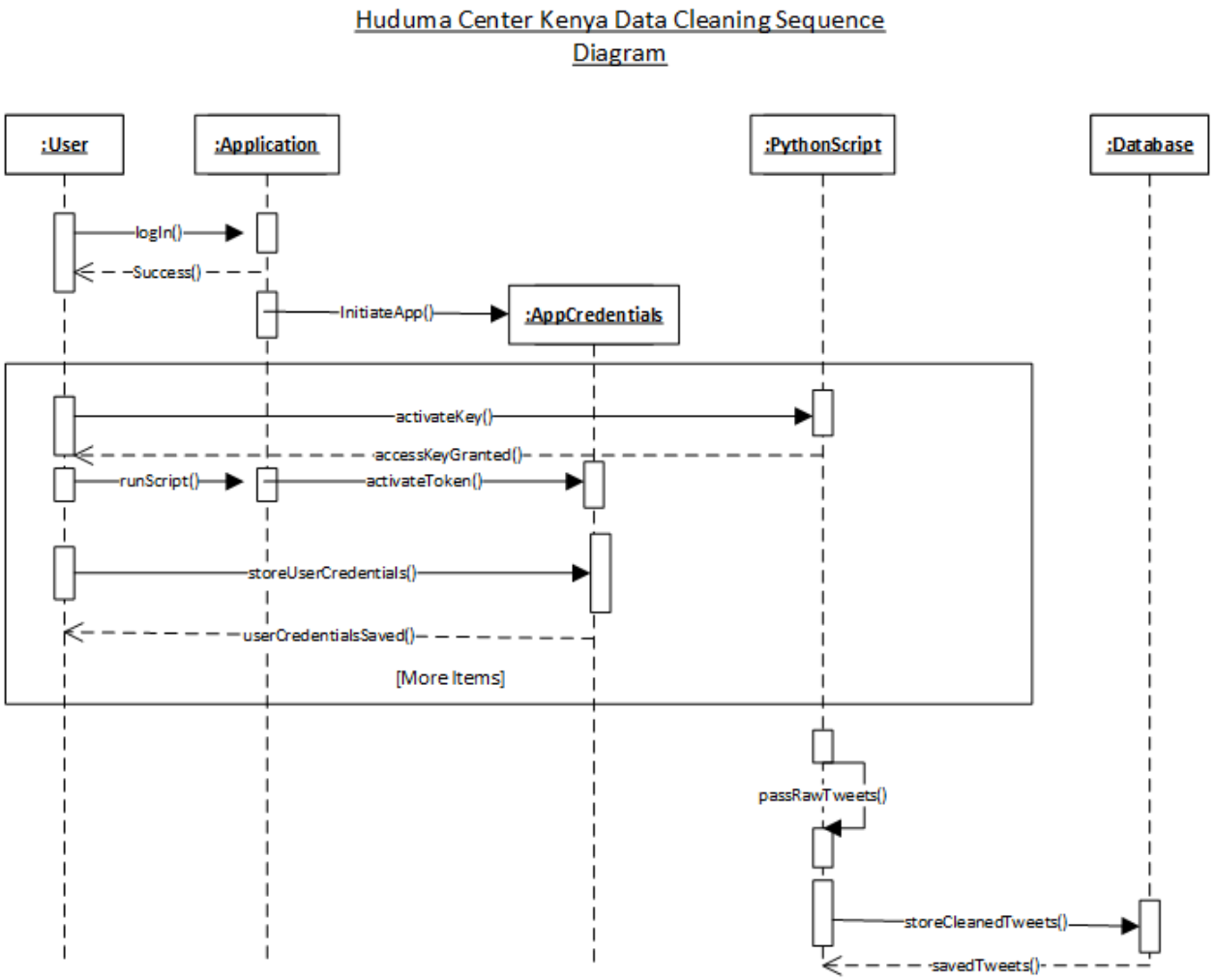


Figure 4.4: Sequence Diagram for Data Cleaning

4.5.4: Sequence of Events

I. The user activates the process

II. Prototype establishes a connection with the database

III. Data is read by the prototype

IV. The prototype script removes unwanted twitter content based on the datasets provided

V. The cleaned data is stored in the database

VI. The prototype terminates

Chapter 5 : Implementation and Testing

5.1: Introduction

This chapter focuses on the implementation and testing of the proposed Mapping of Tweets prototype. The implementation part of this section looks at the different parts of the prototype, how it was developed and how it work. It looks at the tweets being filtered from the Huduma Kenya twitter handle users and the database storage of the tweets filtered for administrative purposes. The testing section looks at functionality and usability testing, ensuring that the prototype meets all of the requirements and objectives that have been established in chapter 1.

5.2: Implementation

The implementation has a brief description of the platform that was used in developing the prototype. This includes principle libraries, environments, and the software language.

5.2.1: Platform Requirements

During the prototype development, a computer system with Pentium processor or higher, 32 bit Intel or higher that is running on windows operating system and a virtual machine that runs on Linux operating system can be used. The current prototype was developed using a 64 bit Intel processor.

5.2.2: Python

Python 2.7.6 was used to implement the prototype due to ease of access and it being open source. It has a variety of libraries, a command interpreter that allows a user to write code, debug code

fast, since no compilation is required. Python has an inbuilt data type for strings, list and more which makes it a robust programming language.

5.2.3: SQLite3

SQLite3 is an open source database that is incorporated in python libraries. It was used as a Database Management System (DBMS) during the prototype implementation since it allows read/write operations and segmentation of the data storage while supporting multithread applications.

5.2.3.1 User Login

The data collection was achieved using the Twitter public streaming API. It is an efficient method of gathering information that is mined from twitter as it allows global access to tweets filtered using various keywords. The author installed python interface library to connect with twitter streaming API. The twitter API has requirements that all its users register an account and provide authentication details when fetching data from the twitter API. Figure 5.1 shows the authentication key used.

```
15  app_k = "bgCDn3oOccJ4giBk2Epd10kPA"
16  app_sc = "TBm4B4mn3MRYTsZLRLKwFHia5WPX94JD84zkqtHbPAx9iBdO3N"
17  OAUTH_TOKEN = "846369573674913793-n7ET933xiufmLvlSVZhDMTB8lyT7MfW"
18  OAUTH_TOKEN_SECRET = "j9Ya5M8jw5IQYwX739B8bdqk3uQ0F4iaqCnuyGLMqXlpB"
19  twitter = Twython(app_k,app_sc,OAUTH_TOKEN,OAUTH_TOKEN_SECRET)
```

Figure 5.1: Twitter API Authentication Key

A python script was also created and executed to provide a connection to twitter API. Data in the format of JSON was pulled from Twitter API and stored in a SQLite3 database. A sample of the collected tweets is shown in Figure 5.2.

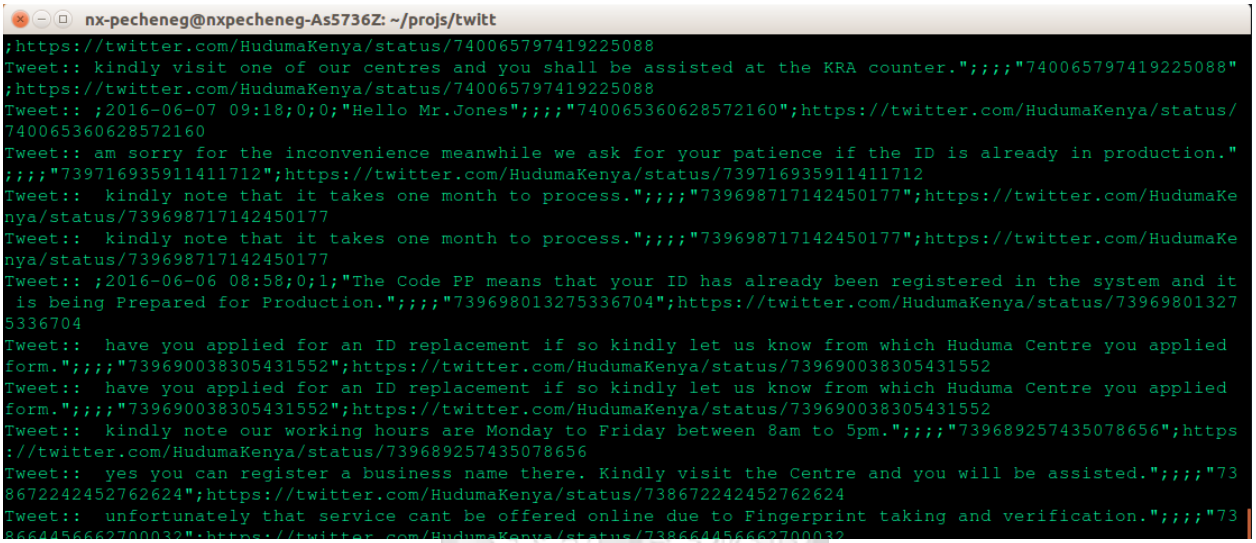


Figure 5.2: Sample of Collected Tweets

5.2.3.2 Word List

This is a list that contains positive and negative words relating to the subject of study which is Huduma Center Kenya tweets for decision making. The word list is stored in a CVS file. The classifier reads this file and associates each tweet with it such that if they are many positive words in a tweet than negative words, it will classify the tweets as positive and vice versa. The word list is shown in Figure 5.3 and Figure 5.4.

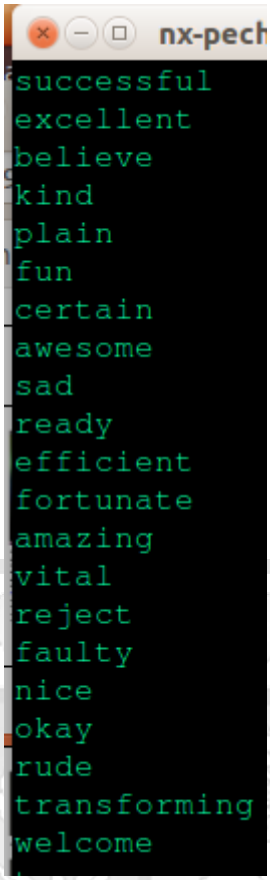


Figure 5.3: Sample of Positive List of Words

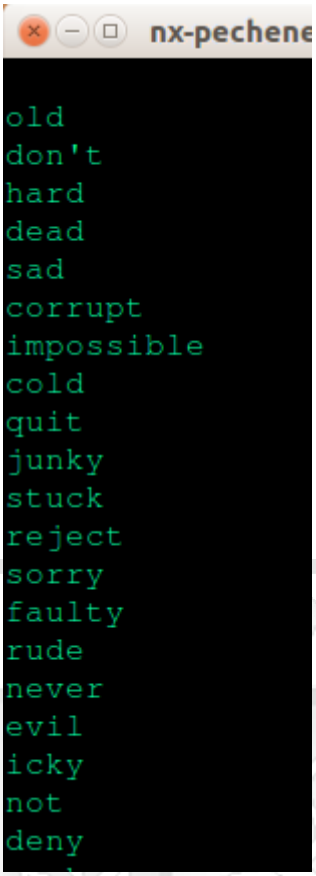


Figure 5.4: Sample of Negative List of Words

5.2.3.3 Data Classification

After the data is collected from twitter, it is classified into two classes either positive or negative tweets using the word list stored in a CVS file as per the sample in Figure 5.3 and Figure 5.4. A tweet was broken down into individual word. Each word was compared to the positive and negative wordlist. If a match was found a counter was incremented depending on the assigned value of word in the wordlist. After completion the classifier saved the results in a text file as shown in Table 5.1 and in Figure 5.6. A python script that performed this action is shown in appendix.

Table 5.1: Sample of Classified Tweets

ID	Tweet	Polarity	Value
13100	this process takes quite some **good** time. Blessed weekend";;;;"847809533737136128";https://twitter.com/HudumaKenya/status/847809533737136128	Positive	2
13113	we will be **glad** to serve you again. Thank you ^D.K";;;;"847774009806774273";https://twitter.com/HudumaKenya/status/847774009806774273	Positive	2
14234	but currently good conduct service is **not** available at Nairobi Huduma Centre. Kindly visit C.I.D Headquarter ^D.K";;@getsh_wifty;;"847036082478833664";https://twitter.com/HudumaKenya/status/847036082478833664	Negative	-1

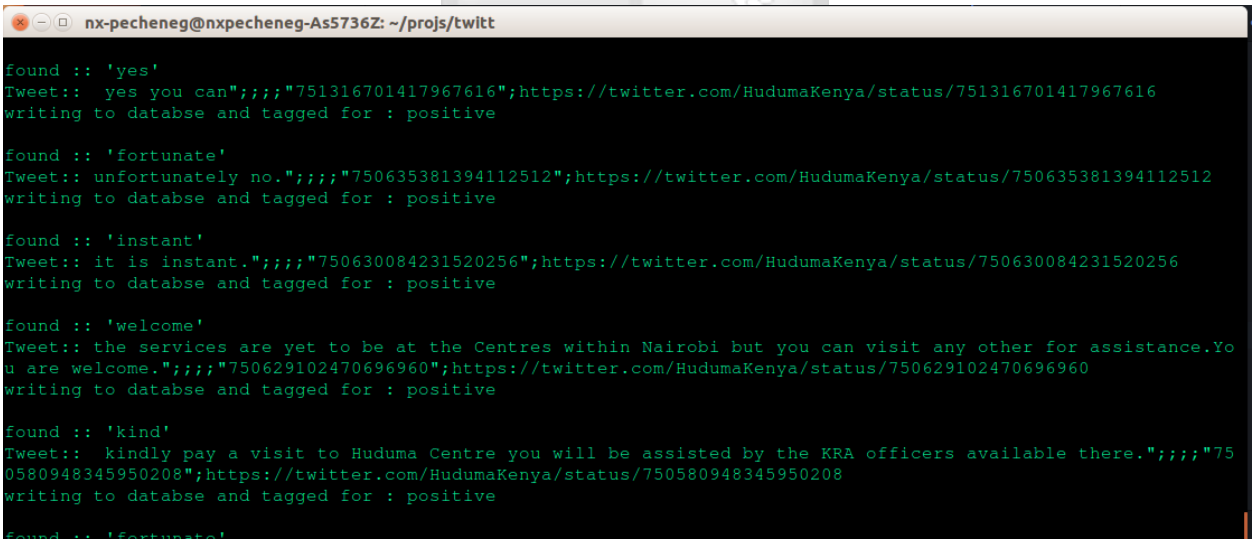


Figure 5.5: Sample of Classified Data

5.3: Testing

5.3.1: Huduma Kenya Polarity

A positive, negative or neutral chart on the tweets is also displayed as in Figure 5.6.

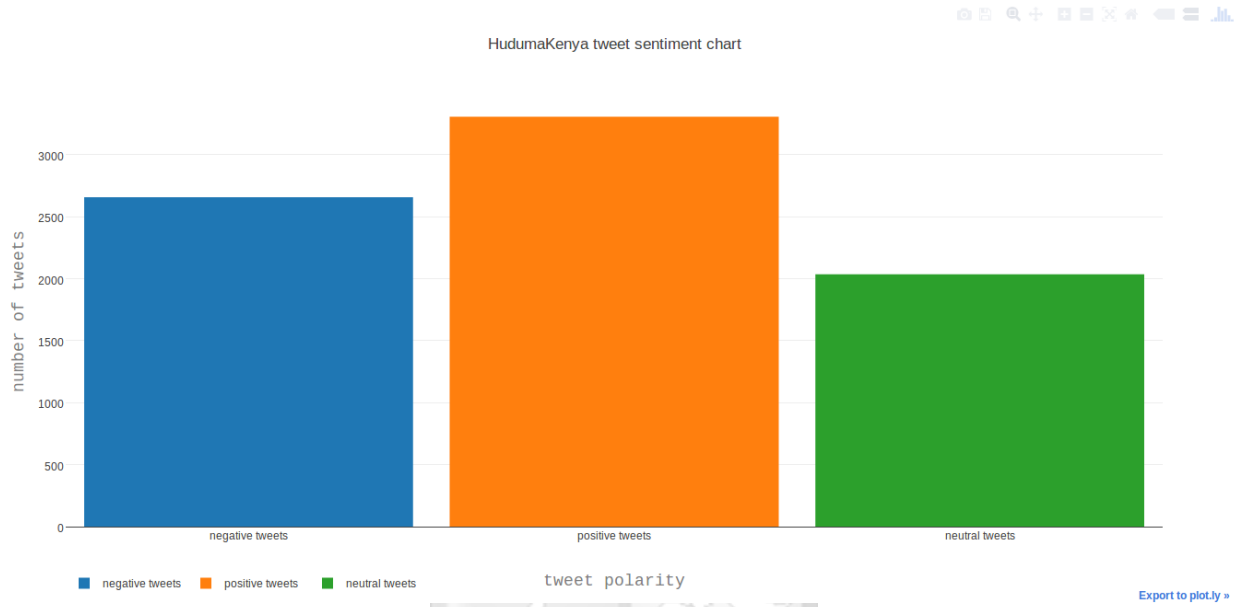


Figure 5.6: Huduma Kenya Sentiment Tweet Chart

5.3.2 Test Results

A table that shows the results of the sample of tweets that were streamed is shown in Table 5.2.

Table 5.2: Test Results

S/No.	Instance	Total
1.	Total number of Instances	8001
2.	Total instances of positive class	3307
3.	Total instances of negative class	2658
4.	Total instances of neutral class	2036

5.3.3 Mapping of Tweets

This is the process of mapping twitter opinions for analyzing to assist in decision making. The map is produced with marks of several locations that are easily to visualize. The three classes are indicated on the map using markers with different colors. The orange color is for positive opinions, the blue color is for negative opinions and the green color is neutral opinions. The map produced by the prototype is shown in Figure 5.7 and Figure 5.8 respectively.

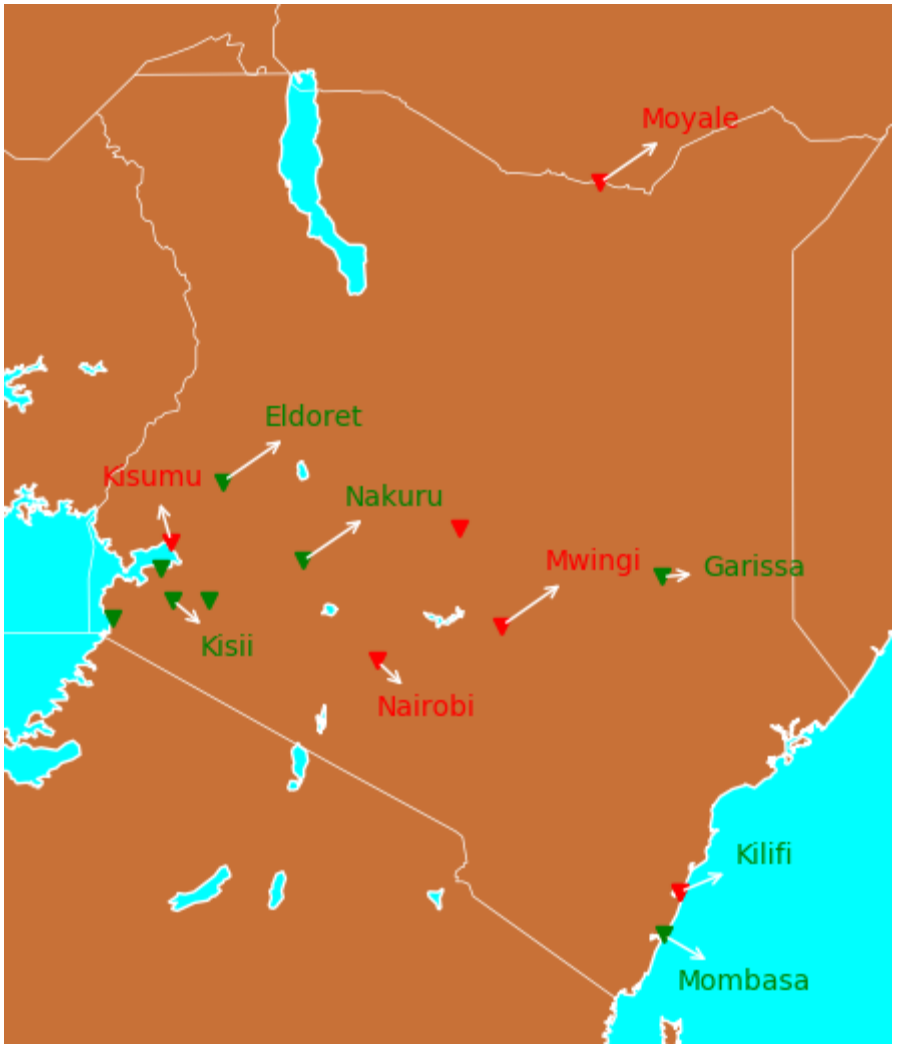


Figure 5.7: Opinion Distribution Map 1

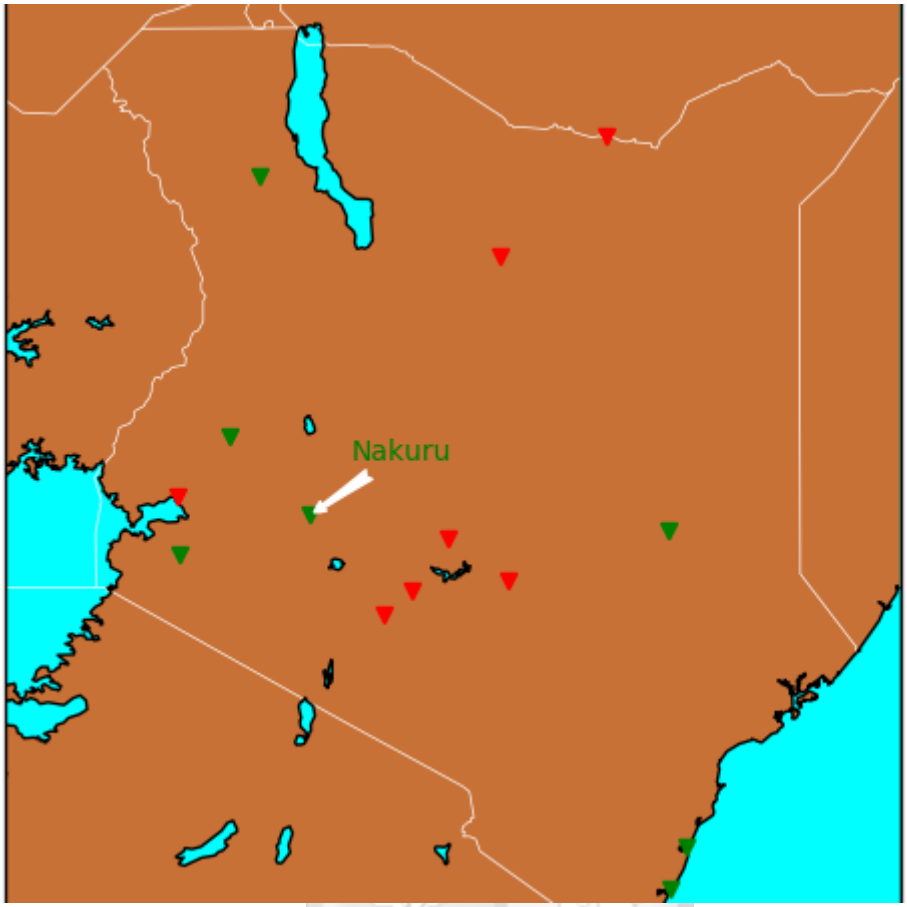


Figure 5.8: Opinion Distribution Map 2

Chapter 6 : Discussion

6.1: Introduction

This chapter reviews the tweets used for the prototype and thus discusses the results and findings obtained from the study. The discussion is based on the objectives, data variables and various relevant literature reviews.

6.2: Prototype Testing

Ease of use and the clarity of the prototype were tested to ensure that the prototype meets the user requirements. The test results from the classifier when evaluated using a dataset containing 8001 tweets collected from Twitter streaming API is displayed in Table 6.1.

Table 6.1: Classifier Results

S/No.	Instances	Total
1	Total No. of Instance	8001
2	Total Instances of positive class	3307
3	Total Instance of negative class	2658
4	Percentage of classified positive instances	41.33%
5	Percentage of classified negative instances	33.22%
6	Percentage of classified neutral instances	22.45%

6.2.1: Visualization

The visualization of the opinions collected from twitter API is displayed and presented by the map in Figure 6.1

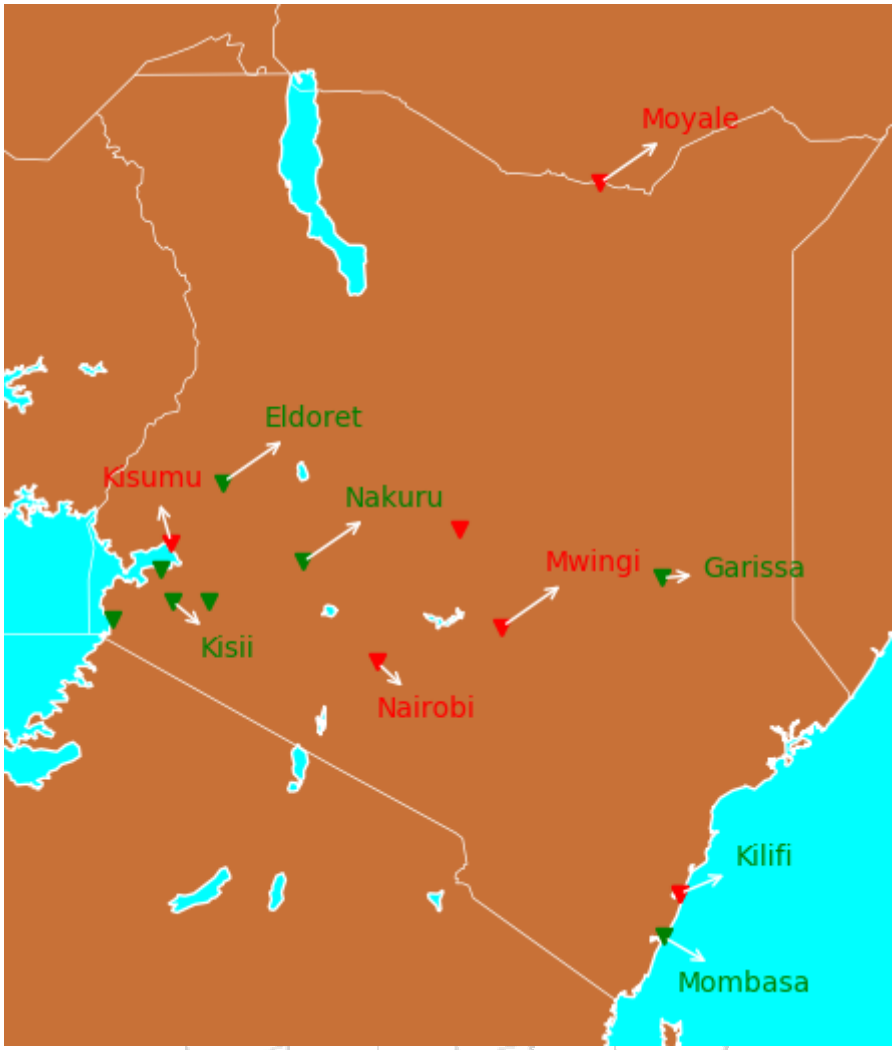


Figure 6.1: Opinion Distribution

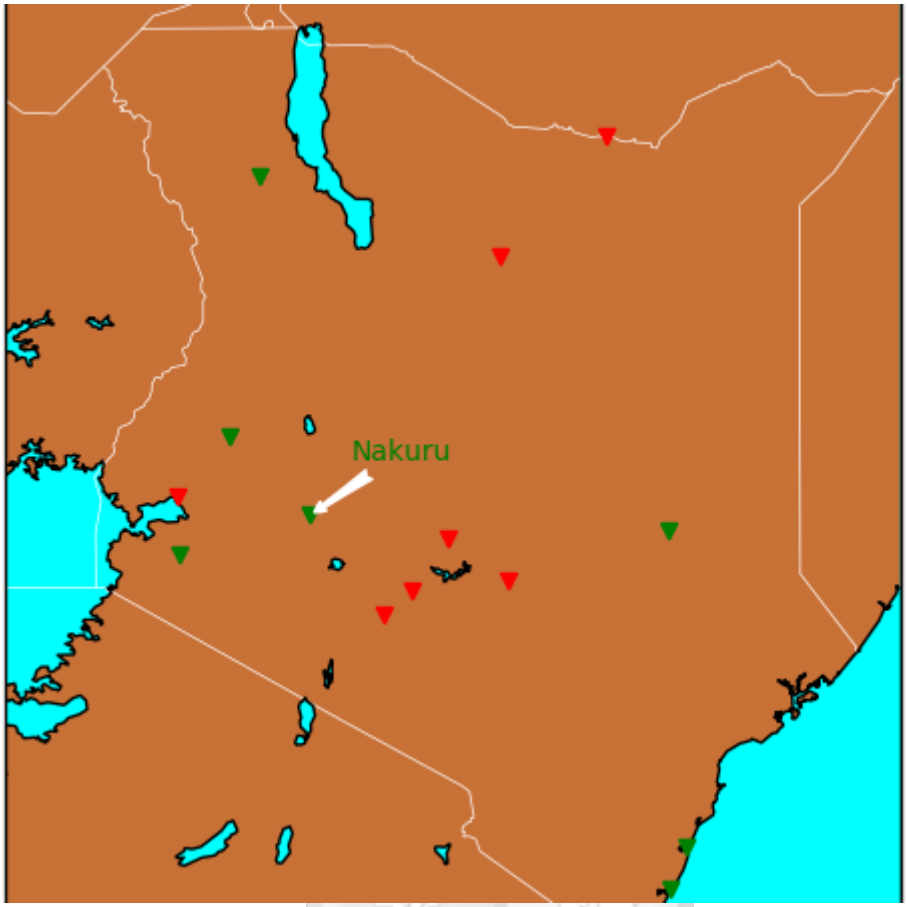


Figure 6.2: Opinion Distribution Mapping

The results display that they are high positive opinions on Huduma Center Kenya than negative opinions. They are also opinions that were classified as both positive and negative making them neutral at a percentage of 22.45%. The positive results are in green color and negative results are in red color as per Figure 6.1 and Figure 6.2.This results can be used by the Kenyan Government when decision making on how efficient and effective are the services being offered at Huduma Centers in Kenya.

6.3 Prototype Implementation

A confusion matrix is a table used to describe classification model performances or classifiers on a set of test data for which the true values are known. In this prototype we use the following equation:

Accuracy = (TP + TN)/ (TP+TN+FP+FN)

Precision = TP/ (TP+FP)

Recall = TP/ (TP+FN)

Whereby, TP is True positive instance. TN is True negative instance, FP is false positive instance and FN is false negative instance. As Doaa (2016) describes it in the Table 6.2.

Table 6.2: Confusion Matrix

N=8001	**Predicted positive**	**Predicted Negative**
Actual Positive	TP=3307	FN=649
Actual Negative	FP=4694	TN=2658

The true positive (TP) rate or recall is the rate at which positive words are predicted to be positive (R), whereas the true negative rate is the rate at which negative words are predicted to be negative. The accuracy represents the rate at which the prototype predicts results correctly (Doaa, 2016). The precision also known as predictive rate, calculates how close the measured values are to each other (P) (Doaa, 2016).

6.4: Performance Evaluation

The overall system performance on accuracy level was 66.78% thus making it more efficient to be used in classifying and mapping twitter opinions for government services that will help

stakeholders in decision making. Table 6.3 captures this information with a 60% precision and 15% recall.

Table 6.3: Overall System Performance

S/No.	Metrics	Accuracy Score
1.	Classification accuracy	66.78%
2.	Precision	60%
3.	Recall	15%

6.5: Challenges

Availability of tweets from all counties in Kenya was a main challenge for collection of co-ordinates. This meant that some missing tweets on location biasness had to be replaced by a null. This was not a true representation of all Huduma Kenya offices countrywide since the offices are all opened but most tweets are from the city center. Time constraints also affected the full development of the system since some external variables could not be incorporated into it.

Chapter 7 : Conclusion and Recommendation

7.1: Introduction

The main objective of this research was to develop a prototype for mapping tweets on state service that will be utilized to assist in decision support by Huduma Kenya stakeholders. This objective was split into five major objectives in order to adequately achieve the main objective. These major objectives are as follow:

i) To establish characteristics associated with mapping of tweets.
ii) To investigate the factors relating to the utilization of Twitter for enterprise decision support.
iii) To appraise the current techniques, approaches and architecture used by enterprises for decision support.

The first three objectives were met by reviewing existing literature. These techniques helped the researcher to identify the characteristics associated with mapping tweets, and enabled the researcher to understand which elements contribute to the final utilization of Twitter for enterprise decision support. This was discussed in chapter 2.

iv) To develop a prototype for mapping of tweets on state services for decision support.

The system was developed using secondary data. Sentimental analysis was used, and this is discussed in chapter 3. The design for the system included the use case, sequence diagrams and a database schema. These designs were used to model how the system components interact, and therefore guide the development of the prototype. This objective was met and is covered in chapter 4 and chapter 5. The tools used in the design of the system included Visio.

v) To test the functionality of the developed prototype.

7.2: Conclusion

The functionality of the developed prototype was evaluated by validating the outputs generated from the developed prototype against the expected output. The effortlessness of use of the prototype was confirmed by use of very simple interfaces. This objective was met and is discussed in chapter 5.

The prototype generated was able to produce predictions that had high accuracy and low errors as indicated by the findings in the study. This illustrates the ability of streamed tweets to be used at one-stop centers such as Huduma Kenya for topic discovery and decision support systems. The study findings also pointed out the need to have more variation of words in the wordlist which can improve the results as per the user requirements on the type of feedback they are looking for from the tweets. The prototype accuracy can also be improved by having more conditions and rules that cover these wordlist. This would help in the monitoring and follow-up of the queries to ensure appropriate state services are provided.

7.3 Recommendations

From the results obtained, the following recommendations can be made:

i) The workflow and tools used in our analysis can be of great benefit by upcoming scholars who wish to draw meaningful insight from aggregate information streamed from twitter.

ii) Government expert can integrate the study of huge data from twitter in their decision support systems.

7.4: Suggestions for Future Research

Upcoming scholars may wish to look into various ways of accurately predictive power of existing techniques on tweets to deliver better geo- locations and appropriate answers to possible twitter queries.

The prototype can be enhanced further to be an entire system with more visible features for example: the user tweeting can see if their comment was implemented and how it has helped other Huduma Center clients. This will improve the visibility functionality of the prototype overall.

Forthcoming scholars may wish to include the analysis to be continental and worldwide to provide valuable insights about Huduma Kenya e-services across the continent or worldwide.

References

Algorithm (n.d.). *Merriam-Webster*. Retrieved March 15, 2016, from http://www.merriam-webster.com/dictionary/algorithm

Abrahams, A. S., Jiao, J., Wang, G. A. & Fan, W. (2012). "Vehicle Defect Discovery from Social Media," *Decision Support Systems,* 54/1 (December 2012): 87-9712

Aggarwal, C. C. (Ed). (2007). Data Stream models and algorithms. NY. U.S.A. IBM. T.J. Watson Research Center Yorktown Heights

Azad, B., Faraj, S., Kane, G., & Majchrzak, A. (2013). The contradictory influence of social media affordance on online communal knowledge sharing. *Journal of computer mediated communication, 19,* 38–55. doi: 10.1111/jcc4.12030

Bailey, B. P. & Horvitz, E. (2010). What's your idea? A case study of a grassroots innovation pipeline within a large software company. *In Proceedings of the 2010 International Conference on Human Factors in Computing Systems* (pp 2065 -2074). NY: ACM

Begelman, G. G. (2006). Automated tag clustering: Improving search and exploration in the tag space. In Proceed of the Collaborative Web Tagging Workshop at WWW06

Bhattacharya, D., & Mitra, M. (2013). Analytics on big fast data using real time stream data processing architecture. EMC proven professional sharing knowledge

Bizhanova, A., & Uchida, O. (2014). Product Reputation Trend Extraction from Twitter. *Social Networking,* 3, 196-202. http://dx.doi.org/10.4236/sn.2014.34024

Borgatti, S., Mehra, A., Brass, D., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science, 323*(5916), 892-895. Retrieved from http://www.jstor.org.ezproxy.library.strathmore.edu/stable/20403062

Boyd, D. M., & Ellison, N. B. (2008). Social network sites: Definition, history and scholarship. *Journal of Computer Mediated Communication, 13*(1), 210-230.

Central Intelligence Agency (C.I.A). Retrieved October 27, 2015 from https://www.cia.gov/library/publications/the-world-factbook/fields/2116.html

Cachia, R. , Compañó, R. & Da Costa, O. (2007). "Grasping the Potential of Online Social Networks for Foresight," *Technological Forecasting and Social Change*, 74/8 (October 2007): 1179-1203.

Cannon, S., & Summers, L. H. (2014). How Uber and the Sharing Economy Can Win Over Regulators. Retrieved April 27, 2016.

Cisco Systems. (2014). Cisco Connected World Technology Report. Retrieved September 2, 2015 from http://www.cisco.com/c/en/us/solutions/enterprise/connected-world-technology-report/index.html

Cooper, R. D. & Schindler, S. P. (2014). Business research methods. McGraw-Hill Irwin Incorporation, New York, NY.

El-Din, D. M., (2016). Enhancement Bag-of-Words, Model for solving the challenges of sentiment Analysis, (IJACSA) International Journal of Advanced Computer Science and Application, Vol. 7.

Deitrick, W & Hu, Wei (2013). Mutually Enhancing Community Detection and Sentiment Analysis on Twitter Networks. *Journal of Data Analysis and Information Processing*, 1, 19-29. http://dx.doi.org/10.4236/jdaip.2013.13004

Fresno, V., Spina, D., Mart´ınez, R., & Zubiaga, A. (2013). Real time classification of tweet trends. *Journal of the American Society for Information Science and Technology*

Folorunso, O. Vincent, R.O., Adekoya, A.F. & Adewal, O.O. (n.d). Diffusion of Innovation in Social Networking Sites among university students. In International Journal of Computer Science and Security (JCSS), (4) 3, pp. 361-372.

Falkowski, T., Barth, A.& Spiliopoulou, M. (2008) Studying Community Dynamics with an Incremental Graph Mining Algorithm, *Proceedings of the* 14*th Americas Conference on Information Systems* (*AMCIS*), Toronto, 14-17 pp. 1-11.

Kailer, D., Kanmaz, K., Lackermair, G., (2013). Importance of Online Product Reviews from a Consumer's Perspective. *Advances in Economics and Business*. 1(1), 1-5. doi: 10.13189/aeb.2013.010101

Doorn, J. van, Lemon, K. N., Mittal, V., Nass, S., Pick, D., Pirner, P., Verhoef, P.C., (2010). Customer Engagement Behavior: Theoretical Foundations and Research Directions. *Journal of Service Research* 13, 253–266.

Empircal. User experience design and Research. Retrieved October 30, 2015 from http://empiricalux.com/?page_id=26

Gartner Research. Gartner Research Methodologies. Retrieved July 22, 2015 from http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

Gemmell, J., Shepitsen, A.,. Mobasher, B. & Burke, R. (2008). Personalizing navigation in folksonomies using hierarchical tag clustering. Proceedings of DaWaK, 196–205,

Girvan M & Newman, M. E. J. (2002) Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12).

Houaiss, B. A "Dicionário Houaiss da Língua Portuguesa," Houaiss Portuguese Dictionary, 2001.

Hsu, D. (2015). In Wharton University of Pennsylvania website. Retrieved on October 25 2015 from http://knowledge.wharton.upenn.edu/article/how-a-detour-can-help-start-up-success/

Jouret, G. (2009). Inside Cisco's search for the next big idea. *Harvard Business Review*, 87(9), 43 – 45.

Kalia, G. (2013). Chitkara University, Punjab. A Research Paper on Social media: An Innovative Educational Tool. *Issues and Ideas in Education* (IIE), Volume 1, March 2013. pp. 43–50

Kaplan, A. M. , & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons, 53*(1), 59-68.

Kaplan, A. M., & Haenlein, M. (2011). The early bird catches the news: nine things you should know about micro-blogging. Business Horizons, 54, 105–113.

Kim, A. J. & Ko, E. (2012). "Do Social Media Marketing Activities Enhance Customer Equity? An Empirical Study of Luxury Fashion Brand," *Journal of Business Research*, 65/10 (October 2012): 1480-1486.

Kumar, R. (2011). Research Methodology a step-by-step guide for beginners. 3rd Ed. Sage, Los Ageles, CA.

Organisation for Economic Co-operation and Development (OECD) (2013). Reliability of public services: Ensuring citizens' rights", in Government at a Glance 2013, OECD Publishing

Oyugi, W. (2015). The Performance of the Higher Civil Service on Policy Management

Pak, A. & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Proceedings of the 7th International Conference on Language Resources and Evaluation*, 1320-1326.

Psck. (2014). Public Service Evaluation Report, Government printers.

Zafarani,R. & H. Liu, (2009). Social Computing Data Repository at ASU [http://socialcomputing.asu.edu]. Tempe, AZ: Arizona State University, School of Computing, Informatics and Decision Systems Engineering.

Saxton, G.D., Oh, O., &. Kishore, R. (2013). "Rules of Crowdsourcing: Models, Issues, and Systems of Control," *Information Systems Management*, 30/1 (Winter 2013): 2-20.

Shariati, M., Bahmani, F. & Shams, F. (2011). Enterprise information security, a review of architectures and frameworks from interoperability perspective.

Singh, K. Y. (2006). *Fundamentals of research methodology and statistics*. New age international ltd, New Delhi.

Social Media (2015). In *Merriam-Webster's online dictionary.* Retrieved from http://www.merriam-webster.com/dictionary/socialmedia

Sherwood Applied Business Security Architecture (SABSA) (2015). In SABSA online website. Retrieved from http://www.sabsa.org/

Twitter (PCMag, n.d.). PC Magazine online encyclopedia. Retrieved on March 15, 2016, from http://www.pcmag.com/encyclopedia/term/57880/twitter

Masdarolomoor, Z., Azmi, R., Aliakbary, S. & Riahi. N. (2011). Finding community structure in complex networks using parallel approach. In Embedded and Ubiquitous Computing EUC, IFIP 9th International Conference

McGee, M. (2012). Twitter: 60% of users access via mobile. Retrieved on March 21, 2016 from http://marketingland.com/twitter-60-percentof-users-access-via-mobile-13626.

McGonigal, E. (2016). Mapping Twitter Topics Network. Retrieved on February 9, 2016 from https://emcgonigal2.wordpress.com/2016/04/13/mapping-twitter-topic-networks/

Muller, M., Erlich, K., Matthews, T., Perer, A., Ronen, I., and Guy, I. (2012), Diversity among enterprise online communities*: Collaborating, teaming, and innovating through social media. In Proceedings of the CHI Human Factors in Computing Systems Conference*, Austin, Texas.

Naylor, R., Lamberton, C., & West, P. "Beyond the 'Like' Button: The Impact of Mere Virtual Presence on Brand Evaluations and Purchase Intentions in Social Media Settings," *Journal of Marketing*, 76/6 (November 2012): 105-120.

Nelson, L. , Convertino, G. , Chi, H. , and Nairn. Studying the adoption of Mail2Tag: an enterprise 2.0 tool for sharing. In Proc. of the 12th European Conference on Computer Supported Cooperative Work (ECSCW), Springer (2011).

Ng'aru, S. W & Wafula M. K. (2015). Factors Influencing the Choice of Huduma Centers' Services :A Case Study of Mombasa Huduma Centre. *International Journal of Scientific and Research Publications*, Volume 5, Issue 6,1 ISSN 2250-3153

Tierney, M. L. & Drury, L. (2013), Continuously Improving Innovation Management through Enterprise Social Media. *Journal of Social Media for Organizations, 1(1), 1-5.*

Toyama, K., & Cooper, M. (2012). Internet freedom: Not a foreign-policy issue. *Issues in Science and Technology, 28*(4), 18-20. Retrieved from http://www.jstor.org/stable/43315635

Yeung, C. M., Gibbins, N. & Shadbolt, N. (2009) Contextualizing tags in collaborative tagging systems. Proceedings of ACM Conference on Hypertext and Hypermedia, 251–260 doi10.1145/1557914.1557958

Watson H.J., Sprague R.H.(1993). The components of an architecture for DSS. In *Decision Support Systems: Putting Theory into Practice*, 3rd ed. Prentice-Hall: New York, NY, USA

Wang, C. (2011). Your anti-social behavior stops here. Set guidelines to adopt social media and reduce risk. Cambridge, MA: Forrester Research, Inc.

Appendices

Appendix A: Python Program

Database creation program

```
                def receiveBuffer(tweets):
                        for t in tweets:
                                outputFile.write(('\n%s;%s;%d;%d;"%s";%s;%s;%s;"%s";%s' % (t.username,
:.date.strftime("%Y-%m-%d %H:%M"), t.retweets, t.favorites, t.text, t.geo, t.mentions, t.hashtags, t.id,
:.permalink)))
                        outputFile.flush();
                        print 'More %d saved on file...\n' % len(tweets)

                got.manager.TweetManager.getTweets(tweetCriteria, receiveBuffer)

        except arg:
                print 'Arguments parser error, try -h' + arg
        finally:
                outputFile.close()
                print 'Done. Output file generated "output_got.csv".'

.f __name__ == '__main__':
        main(sys.argv[1:])
```

Mapping program

```
127     def csv_match(self,tweets_str):#search for word match from tweets
128         for item in range(len(self.compare)):
129             term = self.compare[item]
130             try:
131                 indx = tweets_str.index(term)
132             except ValueError as e:
133                 pass
134             else:#sanitize and store to database
135                 #print '\n'
136                 print "found :: %r"%term
137                 self.terms.append(term)#save all terms
138                 print "Tweet::",tweets_str
139                 print "writing to databse and tagged for :",self.tag,'\n'
140                 tw_uni = unicode(tweets_str,errors = 'replace')
141                 db_save = [tw_uni,self.tag]
142                 dd_save1 = tuple(db_save)
143                 self.to_db.append(dd_save1)
144                 self.matches += 1
```

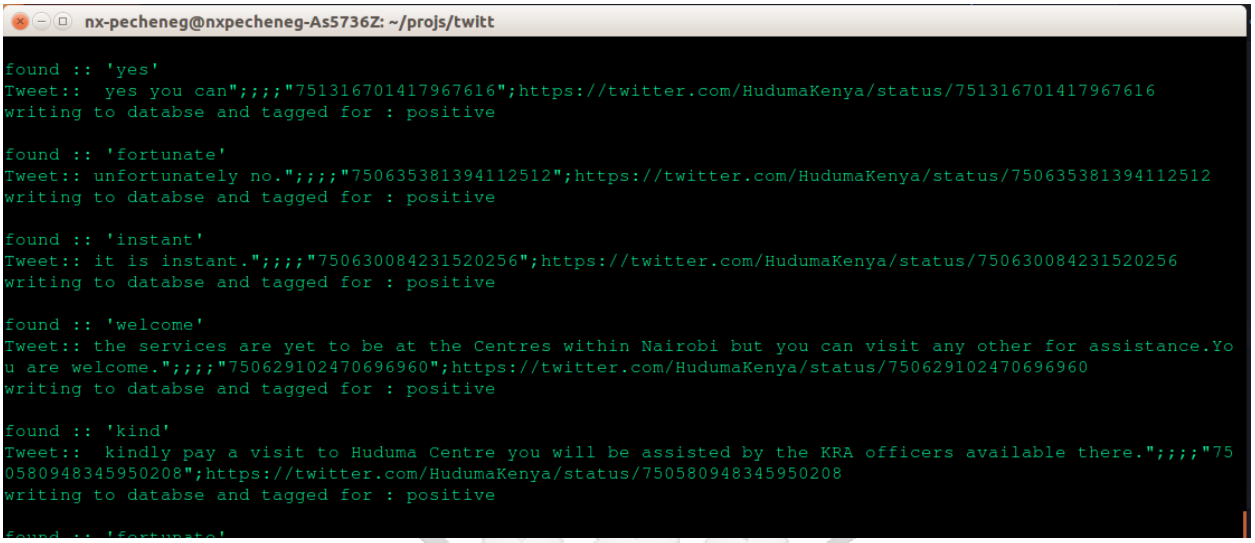
Writing Data to Database

```
            else:
                print "Writing data to Huduma Database"
                conn.executemany('insert into tweet_table values(?,?)',self.to_db)
                conn.commit()
                print "Data save success"
            conn.close()
```

Location Co-ordinates

```
        y_values = [self.neg_tweet,self.pos_tweet]
        x_values =['negative tweets','positive tweets']
        text = ['negative tweets','positive tweets']
        colors = dict(color = ['rgba(255, 0, 0, 0.5)','rgba(0, 255, 0, 0.5)'])
        trace1 = go.Bar(
            x= [x_values[0]],
            y= [y_values[0]],
            #marker=dict(color = ['rgba(255, 0, 0, 0.5)']),
            name='negative tweets')
        trace2 = go.Bar(
            x= [x_values[1]],
            y= [y_values[1]],
            #marker=dict(color = ['rgba(0, 255, 0, 0.5)']),
            name='positive tweets')
        data = [trace1,trace2]
         layout = go.Layout(bargap = 0.1,
                            showlegend = True,
                            legend = dict(orientation = 'h'),
                            title = "HudumaKenya tweet sentiment chart",
                            xaxis=dict(title="tweet polarity",
                            titlefont=dict(family='Courier New, monospace',
                                           size=18,
                                           color='#7f7f7f')
                                           ),
                            yaxis=dict(title='number of tweets',
                            titlefont=dict(family='Courier New, monospace',
                                           size=18,
                                           color='#7f7f7f')
                                           )
                            )
         fig = go.Figure(data=data,layout=layout)
         plotly.offline.plot(fig)
```

Positive Tweet Sample



Negative tweet classification Sample