

# Web 2.0 Computing and Social Media as Solution Enablers for Economic Development in Africa

Patricia E. Nalwoga Lutu

**Abstract** Web 2.0 technologies have been used to create many social media applications that have revolutionalised the way people communicate and socialise. Today, it can be claimed that citizens, especially the young generation, of every country of the world have contributed to the widespread usage of social media. Examples of social media are Facebook, Twitter, LinkedIn and YouTube. Businesses, decision makers in the public sector, news networks and other types of news media have all been quick to realise the potential benefits of social media. Social media analytics are used by many organisations to establish the sentiment of customers towards the products and services that they sell or provide. The purpose of this chapter is to highlight (1) the available and widely used social media and their adoption on the African continent, (2) organisations on the African continent that have adopted these social media, (3) how organisations that use social media can benefit from such usage, and (4) the challenges posed by the adoption and usage of social media. It is argued that, for African countries, the major benefits of using social media are the low costs of engaging with the citizens and business customers as well as the ability to analyse the effectiveness of such engagements.

## 1 Introduction

Social media can be defined as Internet-based media that are created and shared by communities. Web 2.0 technologies can be defined as Internet technologies which connect people and enable the sharing of media. In the last 10 years, Web 2.0 technologies and social media have revolutionalised the way people communicate and socialise. Social networks are one category of social media that facilitates the formation of communities and sharing of content such as microblogging (Twitter), meeting people and sharing content (Facebook, Twitter, LinkedIn, YouTube). Twitter, Facebook, LinkedIn and YouTube have become very popular in many countries including African countries, especially with the young generation.

---

P.E.N. Lutu (✉)

Department of Computer Science, University of Pretoria, Pretoria, South Africa

e-mail: [Patricia.Lutu@up.ac.za](mailto:Patricia.Lutu@up.ac.za)

Organisations in the public and private sectors have been quick to realise the value of social media. At the present time, business, government and non-governmental organisations typically participate in social media. The reasons for participation include assessing the opinions of the public about products and services that the organisations provide, soliciting opinions from the public, and communication and collaboration between stakeholders. Big and small business organisations are routinely using social media for marketing and branding purposes [4, 20, 29]. Many government departments and other public sector organisations use social media for purposes of engaging with citizens (e.g. [26]). Development organisations such as non-government organisations (NGOs) routinely use social media to network with stakeholders and to reach and engage with developing country communities via mobile phones.

Many online tools are available for the analysis of social media data for purposes of assessing the effectiveness of social media usage. Some tools are specific to a particular service (e.g. TwitterSentiment) while others are general purpose tools that can be used to analyse data from different types of social media services [33]. Some analytics tools may be used for free while others require payment for the services. Organisations can use these online tools to build their online presence and grow their organisation's reach online. Organisations are not limited to the use of online tools for social media analytics. In fact, data can be downloaded from a social network website, stored in special databases and then analysed offline using statistical, data mining, and machine learning methods [2]. NoSQL databases are a good candidate for the storage of large quantities of unstructured data. Some of the methods of knowledge discovery and data mining that can be applied to social network data are sentiment analysis, time series analysis, and graph mining [2, 24, 34].

Given the foregoing discussion, it is argued in this chapter that Web 2.0 technologies and social media may be viewed as a key solution enablers for government and NGOs, and private sector organisations in African countries. These technologies have the potential to connect the citizens, public sector organisations, private sector organisations and development organisations on the African continent at a very low cost and in very short time frames. The objectives of this chapter are to highlight (1) the available and widely used social media and their adoption on the African continent, (2) organisations on the African continent that have adopted these social media, (3) how organisations that use social media can benefit from such usage, and (4) the challenges posed by the adoption and usage of social media. The rest of this chapter is organised as follows: Section 2 provides a discussion of Web 2.0 and social media. Section 3 provides a discussion of organisations that use social media. Section 4 discusses social media adoption in African countries. Section 5 discusses online social media analytics tools. Section 6 discusses offline analysis of social media data. Section 7 provides a discussion of the benefits, challenges and some recommendations for social media data analytics to support development. Section 8 concludes the chapter.

## 2 Web 2.0 Technologies and Social Media

Gartner [12] have defined social media as ‘an online environment in which content is created, consumed, promoted, distributed, discovered or shared for purposes that are primarily related to communities and social activities rather than functional task-oriented objectives’. Gartner have also defined Web 2.0 as ‘the evolution of the Web from a collection of hyperlinked content pages to a platform for human collaboration and system development and delivery’ [12]. Web 2.0 technologies and social media have revolutionalised the way people communicate and socialise. This section provides a summarised discussion of Web 2.0 technologies and social media services.

### 2.1 Web 2.0 Technologies

Most commonly, Web 2.0 technologies are defined in terms of what they enable people to do, that is, their main agenda, which is to connect people in numerous ways so that they can utilise their collective strengths [7]. In this chapter, the term Web 2.0 technologies is used to refer to web technologies that enable the creation and sharing of social media. The term social media is used to refer to online (Internet) media that is created and shared by communities. Table 1 provides a summary of some of the well-known categories of social media.

As shown in Table 1, different types of social media enable people to express their opinions and get feedback from others (blogs), express their opinions concisely about what is happening right now (microblogs), meet other people and share content with them (social networks), collaborate to generate content (wikis), and upload and share media content (media sharing).

**Table 1** A sample of categories of social media

Media category	Purpose	Example of service
Blogs (web-logs)	Facilitate the expression of personal opinions by the public	There are many examples of blogs on the web
Microblogs	Facilitate the expression of personal opinions about what is happening right now	Twitter
Social networks	Professional or social networking sites which facilitate meeting people and sharing content	Facebook, LinkedIn, Twitter
Collaborating	Collaborative reference works (e.g. Wikipedia) that are built using wiki-style software tools	Wikipedia
Media sharing	Facilitate the sharing of digital media, e.g. videos	YouTube

## 2.2 Social Networks

Dasgupta and Dasgupta [7] have provided a discussion of different types of social networks that existed in 2009. Social contact networks are primarily used for friends and family, e.g. Facebook, Twitter. Study circles are networks dedicated to students. Social networks for specialist groups are used by core field workers like doctors, engineers, members of corporate industries, e.g. LinkedIn. Police and military networks are private social networks (not in the public domain) exclusively for people in these services. There are other network categories which include sporting networks, networks for fine arts, shopping and utility networks. A brief discussion of Twitter, Facebook and LinkedIn social networking services is given in this section.

### Twitter: A Social Network and Microblogging Service

A number of microblogging services exist on the web, with Twitter [32] possibly being the most visible (popular) at the moment. Twitter is a ‘whats-happening-right-now’ social media service [2]. Twitter was launched in 2006 as a microblogging service that allows users to send updates (tweets) to a network of friends (followers) from a variety of devices [20]. Twitter users need a subscription in order to receive updates and tweets are delivered instantly. Tweets are displayed on the user’s profile page on Twitter, or they can be delivered via instant messaging (SMS—short message service), Really Simple Syndication (RSS), e-mail or through an application such as Twiterrific or Facebook [20]. Tweets are at most 140 characters. The users interact by *following* updates of people who post interesting tweets. Users can pass along interesting pieces of information to their followers. This is known as *retweeting*. Users can also respond to (or comment on) other people’s tweets which is called *mentioning* [6]. The following is an example of a tweet: *RT @toni has a cool #job*. RT is used at the beginning of a tweet to indicate that the message is a *retweet*. Users can reply to (mention) other users by indicating user names prefixed with the @ character (e.g. @toni). Hashtags (#) are used to denote subjects or categories (e.g. #job) [2]. Additionally, emoticons such as smiley (:-), sad face (:-) and variations of these are added to tweets to express sentiment [2, 34]. Tweets may be kept private among the followers, or they may be made public and unrestricted [34].

Twitter is characterised by the large volumes of data that are generated and the large numbers of users. O’Connor et al. [24] have reported that for the 2 year period from 2008 to 2009, the message volume for Twitter increased by a factor of 50. In April 2010, Twitter reported various statistics on the users of Twitter as follows [2, 35]. There were 106 million registered users, 180 million unique visitors every month, and 300,000 new users signing up every day. There were 600 million queries being received daily via Twitter’s search engine and three billion requests per day based on the Twitter application programming interface (API). It was also noted that 37% of active users used mobile phones to send requests. More recently, the number of regular Twitter users has been estimated at 200 million.

### **Facebook: A Social Networking Service**

Facebook was launched in 2005 as a social network for use by university and college students in the USA, and so, it was aimed at the young adult age group (18–24 years old). The inventor of Facebook is Mark Zuckerberg, a former student at Harvard University. Facebook opened its services to non-academic users for the first time in 2007. By 2008, Facebook had grown to be the second largest social network with more than 30 million users [15]. As a social networking site, Facebook facilitates meeting people and sharing content such as photos, blogs, microblogs and Facebook applications developed by the users. Big businesses, small businesses, governments and NGOs have all been quick to realise the benefits of having a Facebook presence. Graham [15] has observed that, from a business perspective, Facebook has given big opportunities for businesses to direct their marketing efforts towards the young adult age group (18–24) who are known to be very difficult to win-over.

### **LinkedIn: A Social Networking Service**

LinkedIn is a business-oriented social networking site. LinkedIn users normally associate with their line-of-work network and use the site to maintain a list of contact details for people (connections) they know and trust within their line of work. The network of contacts is used to maintain communication, exchange trade information, academic information, and other types of information. LinkedIn uses a ‘gated-access-approach’ which means that connecting with other users of LinkedIn requires a pre-existing relationship or the intervention of a mutual contact. This mechanism is designed to create trust among LinkedIn users [25].

### **YouTube: A Media Sharing Service**

YouTube was launched in 2005 and was acquired by Google as a subsidiary in 2006. As stated on the YouTube website ([www.youtube.com](http://www.youtube.com)) the service allows billions of users to discover, watch and share originally created videos. The service provides a forum for people across the globe to connect, inform and inspire others. Video content includes amateur and professional video clips, television clips, music videos, educational videos, and corporate videos. Unregistered users can view videos and registered users can view and also upload videos. According to YouTube Statistics [36], more than one billion unique users visit YouTube every month and millions of new subscriptions happen everyday. YouTube is localised in 61 countries and languages, and 80% of the YouTube usage traffic originates from outside the USA. Mobile devices make up almost 40% of YouTube’s global watch time. Individuals, government organisations, small and big business organisations around the world are using YouTube to grow their audiences.

### 3 Organisations that use Social Media

Organisations have been quick to realise the value of social media. At the present time, business, government and non-governmental organisations typically participate in social media. The reasons for participation include assessing the opinions of the public about products and services that the organisations provide (e.g. Twitter microblogs), soliciting opinions from the public and communication and collaboration between stakeholders (e.g. social networks). This section provides a discussion on how organisations around the world are using social media.

#### 3.1 *Big Business Organisations and Social Media Marketing*

Big business organisations are routinely using social media for marketing and branding purposes. Stelzner [29] has reported the results of a survey on the usage of social media for marketing purposes. Over 3,000 business organisations worldwide, predominantly in the USA, Canada, UK, Australia and India, participated in the survey. One major conclusion from this survey was that, in these countries, the top five social media platforms for marketing are Facebook, Twitter, LinkedIn, blogging, and YouTube, in that order. According to Stelzner [29], 92% of the organisations used Facebook, 80% used Twitter, 70% used LinkedIn, 58% used blogging, and 56% used YouTube. In terms of usage, Stelzner [29] has observed that (business) organisations conduct the following activities for social media: content creation, analytics, monitoring, obtaining status updates (tweets), research, strategy formulation, and community engagement.

Jansen et al. [20] have conducted research to assess the effectiveness and trends of the use microblogging by businesses for purposes of word-of-mouth branding. They specifically studied the use of Twitter for online word-of-mouth branding. They have concluded that (1) microblogging can be used to provide information to customers and the public in general and to draw potential customers to other online media for the business, such as websites and blogs. So, monitoring microblogging sites concerning a business brand and competitors' brands can provide valuable competitive intelligence information. (2) Using microblog monitoring tools, businesses can track postings and immediately intervene with unsatisfied customers. (3) By setting up corporate accounts on microblogging services, businesses can use microblog polls and surveys to obtain near real-time feedback from customers. (4) Businesses can also obtain valuable product improvement ideas by tracking microblog postings. (5) Businesses can take advantage of contacts made via microblogging services to further their branding efforts by responding to comments made about the company brand.

### ***3.2 Small Businesses and Social Media Marketing***

Bodnar [4] has observed that successful small businesses have long thrived on word-of-mouth advertising by satisfied customers in order to help promote their products or services. Due to the ubiquitous nature of the social media services available today, small businesses are in a position to use free tools to help increase word-of-mouth advertising and to decrease the need for expensive advertising channels such as magazines, newspapers, radio and television. Bodnar [4] has reported the results of a study that was conducted on small businesses that have successfully used social media marketing in the USA. Bodnar [4] has also identified a number of challenges and key success factors for small businesses to make effective use of social media marketing. While big businesses have marketing departments that can attend to the time-intensive social media marketing activities, this is not the case for small businesses. However, small businesses can commit weekly resources to creating content and engaging in social media such as Facebook, Twitter or blogs. It is also important for the small business to have some method of establishing how a given social media activity has impacted the business results and to use this information to drive business strategy.

### ***3.3 Public Sector Organisations***

Many government departments and other public sector organisations can use social media for purposes of engaging with citizens and conducting polls for various purposes. O'Connor et al. [24] have conducted studies to compare the results of traditional polling with the results of polls conducted via social media, namely Twitter. Using time series analysis, their studies compared the results of public opinion polls (in the USA) on consumer confidence about the US economy with the results of rudimentary (simple) sentiment analysis of Twitter data on these topics over the same time period. They also compared the results of public opinion polls on the popularity of the US president job approval with the results of sentiment analysis of Twitter data on this topic over the same time period. They concluded that a relatively simple sentiment detector based on Twitter data replicates the results obtained using formal and more expensive polling methods. The findings by O'Connor et al. [24] are obviously good news, especially for developing economies. Decision makers should be able to set up polls and surveys, at very low cost, using microblogging services, in order to engage with the country's citizens, assess public sentiment about the economy and other services provided by government departments and agencies. Public safety organisations have also realised the value of microblogging. Jansen et al. [20] have observed that Twitter is increasingly being used by these organisations to receive updates during emergencies and natural disasters so that they can make informed decisions on how to plan rescue operations.

### ***3.4 Non-government Organisations***

Non-government Organisations (NGOs) are organisations that are involved in development activities, mostly in developing countries. These organisations consist of a number of stakeholders including: donors, fellow NGOs, staff, local organisations in the developing regions and aid-receiving communities [28]. Sheombar [28] has discussed various opportunities that social media presents for NGOs. These include: collaboration, connecting and interacting, networking, international co-operation, and communicating with the aid-receiving communities via mobile phones. Sheombar [28] has reported the results of a study conducted on Dutch NGOs and their usage of social media and has identified some of the benefits and challenges posed by social media usage. When an NGO has knowledge of the local context, they can specifically target certain groups for purposes of information delivery and data collection. Sheombar [28] has reported that collection of data via mobile phones is a wide spread practice in many development organisations (NGOs). The major benefit here is that organisations are able to reach developing country communities that do not have access to computers. It should be noted that collection of mobile data via mobile phones is not the only data collection method employed by NGOs. Specific crowd funding and fundraising is another activity that has been conducted by NGOs via social media. Some NGOs also continually monitor and analyse their social media activities. Two main challenges identified by the organisations surveyed by Sheombar [28] are the need to respond fast on social networks, and the difficulty of conveying complex messages via social media.

## **4 Social Media Adoption in African Countries**

In the last few years, African countries have experienced a widespread adoption of social media, especially with the young generation. It has been reported in the literature that the key drivers for this adoption have been the widespread adoption of mobile phones, establishment of mobile Internet infrastructure on the continent, and the affordability of mobile Internet services for the ordinary person. This section provides a brief discussion of the drivers for social media adoption and the extent of this adoption.

### ***4.1 Drivers for Social Media Adoption***

In August 2012, Deloitte ([www.deloitte.co.za](http://www.deloitte.co.za)) and Frontier Advisory ([www.frontieradvisory.com](http://www.frontieradvisory.com)) hosted an African Frontiers Forum to evaluate the economic impact of social media in Africa. They have reported that in 2012, the African continent was the second largest mobile phone market (after Asia) with more than



700 million mobile connections. Estimations of the annual growth were 30% so that by 2016, the number of mobile connections should rise to almost one billion. They have also reported that the widespread adoption of mobile phones and the roll-out of mobile Internet infrastructure in many African countries has resulted in the availability of affordable Internet services to the vast majority of African citizens. This is in stark contrast to Internet access via fixed line telecoms services which are generally unavailable and unaffordable to the vast majority of citizens. Most newly activated mobile devices are Wireless Application Protocol (WAP) enabled. In 2012, Africa's mobile data usage amounted to 14.85% of the total global Internet traffic. It has been argued that the wide adoption of mobile devices, ease of access to Internet services, and affordability of Internet services via mobile devices have been the major driving factors for social media adoption on the African continent [9].

## ***4.2 Social Media Adoption***

There has been a widespread adoption of social media in recent years, and Facebook has become the most visited website on the African continent. It was reported in 2012 [9] that for the African continent, the users of the Facebook website were estimated at 44.9 million people. Fuseware and World Wide Worx [11] have reported that in 2014, there are 9.4 million active users of Facebook and 5.5 million users of Twitter in South Africa. It has also been reported that the majority of Facebook and Twitter logins (approximately 80%) from Nigeria and South Africa are from mobile devices [9, 11]. The popularity of Facebook in Africa has prompted Facebook to specifically cater to African markets by starting to roll-out local language versions of the website, starting with Swahili. The Swahili language originates from the East African coast (Kenya and Tanzania) and is spoken widely in East Africa and Central Africa. In addition to the big (American-based) social media services, local social media services have come into existence in some African countries. One such example is Mxit in South Africa (<http://get.mxit.com>), a social networking and instant messaging service with an estimated user base of six million subscribers [11].

In the private sector, African businesses are increasingly employing social media strategies to engage more effectively with consumers through continuous interaction and engagement [9, 11]. Fuseware and World Wide Worx [11] have reported that, in South Africa, 93% of South African corporations that are major brands use Facebook, 79% use Twitter, 58% use YouTube and 46% use LinkedIn for marketing and branding purposes. However, less than 10% use the home-grown Mxit service. Fuseware and World Wide Worx have further reported that a survey of South Africa's top 50 brands revealed that, on average, they each have 58,000 Facebook fans, 259,000 YouTube account views and 12,785 Twitter followers. In the public sector, many organisations are also using social media to engage with the public. Many government organisations in developing countries have a Facebook

and Twitter presence. As an example, in South Africa, the office of the Presidency [26], the South African government [13, 14] and the National Department of Health [17] all have a presence on Facebook and Twitter. Many public sector and higher education institutions in Africa also have a Facebook and Twitter presence.

## 5 Online Social Media Analytics Tools

The use of analytics tools is essential for organisations (or individuals) that are serious about building their online presence and growing their organisation's reach online. Many online tools are available for the analysis of social media data. Some tools are specific to a particular service (e.g. Twitter) while others are general purpose tools that can be used to analyse data from different types of social media services. This section briefly discusses some of the available online tools. It should however be noted that online analysis tools tend to come and go rather quickly.

### 5.1 *Online Tools for Analysis of Twitter Data*

The Summize tool was a popular online service for searching tweets and keeping up with emerging trends in real time. This tool was acquired by Twitter in 2008 [20]. Summize enabled users to submit queries requesting for the retrieval of tweets on a given topic or brand for a specified time period, followed by analysis of the sentiments expressed in the tweets. Summize would analyse the sentiment and give an overall sentiment rating using a five-point Likert scale with levels (from lowest to highest) wretched, bad, so-so, swell, and great. Twitter Sentiment (now Sentiment140) is an online tool provided by Twitter for online analysis of tweets. Currently (in March 2014), this tool is available at the website <http://www.tweetersentiment.appspot.com>. The tool enables visitors to this site to research and track the sentiment for a brand, product or topic of interest. This website enables a visitor to track queries over time. A visitor can also retrieve sentiment counts over time and retrieve tweets along with their classification. An API is also provided for sentiment analysis [2]. There are also Web tools for searching for tweets. A sample of such tools is given in Table 2. A detailed list is available from [5]. It should be noted again that web tools appear and disappear very often and very quickly.

### 5.2 *Online Tools for Analysis of Different Types of Social Media Data*

VentureBeat [33] has provided a brief description of the top ten tools for social media analytics that were in use by organisations at the end of 2013. Three of these

**Table 2** Examples of web tools for searching for tweets

Name and URL	What the tool provides
Backtweets, <a href="http://backtweets.com">http://backtweets.com</a>	See how many people have tweeted a specific link
NearbyTweets, <a href="http://nearbytweets.com">http://nearbytweets.com</a>	Tweets by Twitter users near you (uses Google Maps)
Twellow, <a href="http://twellow.com">http://twellow.com</a>	Organise tweets in yellow page format, a Twitter directory
Topsy, <a href="http://topsy.com">http://topsy.com</a>	Springboard of more than 30 million users and what they are talking about in real time

tools are briefly discussed here to give an idea of the functionality provided by these tools. Google Analytics is a free resource for social media analytics on an organisation's (or individual's) website. According to VentureBeat [33], Google added *Social Reports* to analytics in 2012. Organisations can use social reports to determine the conversion value of visitors from social sites as well as see how visitors from different social sites behave on the organisation's website. *Social Reports* also has an activity stream that shows in real time how people are talking about the organisation's website on social networks. *Brandwatch* is a tool that monitors all conversations across various social networks. This tool also supports 25 languages. *Hootsuite* is an analytics tool that offers a single online dashboard that an organisation can use to manage their social media accounts such as Twitter, Facebook, Google<sup>+</sup>, LinkedIn and other accounts. Additionally, tools are provided for social media analytics.

## 6 Offline Analysis of Social Media Data

Organisations are not limited to the use of online tools for social media analytics. In fact, data can be downloaded from a social network website stored in a special database and then analysed offline using statistical, data mining, and machine learning methods [2]. NoSQL databases are a good candidate for the storage of large quantities of unstructured data such as textual data that is stored by social media services. This section provides a discussion of NoSQL databases and methods that have been reported in the literature for the analysis of social media data.

### 6.1 Big Data and NoSQL Databases

Social media services have resulted in huge amounts of unstructured data being generated on a continuous basis, as indicated in Section 2. This data is commonly called Big Data. Big data is defined as data with the following characteristics: big volume, big velocity and big variety. Big volume means that the generated data is at scale of terabytes to petabytes. Big velocity means that the data is in motion,

that is, it is arriving at high speed. Big variety means that the generated data is in many forms, that is, structured, semi-structured, unstructured, text, and multimedia data. Web-generated big data is stored in NoSQL databases because relational database systems are not suitable for storing Big Data. NoSQL database systems are distributed non-relational databases designed for large-scale data storage and for massively parallel data processing using a large number of low cost servers in order to provide scalability, availability and fault tolerance [23]. NoSQL databases arose alongside major Internet businesses which had challenges in storing and processing huge quantities of data. Examples of these organisations are Google, Amazon, Facebook and Yahoo! Today they are used by organisations that collect large amounts of unstructured data for analysis purposes. There are currently four categories of NoSQL databases namely: key-value stores, document stores, wide-column stores, and graph databases [23].

Key-value stores store data entries as key-value pairs where the key uniquely identifies the value (data item). The value may be a word, number or complex structure with unique semantics. Document stores (databases) were inspired by Lotus Notes and are designed to store documents. The documents are encoded in a standard data exchange format, e.g. XML, JSON (Java Object Notation), BSON (Binary JSON). The data is stored in key-value pair style but the value column is unstructured data (document). Primary uses of document stores are storing text documents, e-mail messages and XML documents. Two examples of document stores are MongoDB and Apache's CouchDB [23]. Wide-column stores use a distributed, column-oriented data structure which accommodates multiple values per key. These databases use Google's Bigtable structure and file systems (GFS) and MapReduce parallel processing [23]. Graph databases use structured relational graphs of interconnected key-value pairings. A graph is represented as an object-oriented network of nodes (objects), edges (node relationships), and properties (object attributes expressed as key-value pairings). Graph database systems provide visual representation of information as well as an API for querying the graph data. Primary uses of graph databases include representing social networks, generating recommendations and conducting forensic investigations. Examples of graph databases are Neo4j, InfoGrid and AllegroGraph [23].

Organisations that plan to collect and store social media data from social media services should consider investing in NoSQL database systems for storing this data. Document databases (e.g. MongoDB and Apache's CouchDB), as well as graph databases (e.g. Neo4j) are especially suitable for storing social network data. Additionally, MongoDB, Apache's CouchDB and Neo4j are Free/libre/open source software (FLOSS) databases. FLOSS is software that is licensed to grant users the right to use, copy, study, change and improve its design through the availability of its source code. 'Free' refers to the freedom to copy and re-use the software, rather than to the price of the software [10]. Well-known FLOSS projects include Apache web server, GNU Linux, FreeBSD, MySQL, OpenOffice.org and Mozilla. FLOSS offers a number of benefits for organisations. These include reduced software costs, vendor independence and open standards. For developing countries specifically, FLOSS also eliminates the high costs of dollar-based software licences, since

FLOSS licences are much cheaper and are not specific to a machine. There are a number of recognised challenges associated with FLOSS usage. One is the fact that skills are scarce and therefore more expensive. A second challenge is that there is no accountability or possible recourse to legal claims should there be a major problem with the software. A third challenge is the lack of a 24/7 help desk. These challenges indicate that an organisation must weigh the pros and cons of FLOSS before deciding to adopt it for mission critical applications. If the analysis of social network data is not mission critical (and most commonly it will not be) for an organisation, FLOSS database systems should be seriously considered as a viable and affordable solution for data storage.

## ***6.2 Obtaining Data from Social Media Applications***

An API is a library of class definitions and functions that enable software developers to access and use the low-level functionality of a given system without having to access the source code. Internet-based companies such as Google, Amazon and Yahoo provide APIs for software developers. Social media service providers such as Twitter, Facebook and LinkedIn also provide APIs that enable developers to develop applications that can access data stored by the service, filter and analyse the data in various ways, and enable other users of the service to use the application. Some online analysis services also provide APIs. Some authors have observed that the use of APIs for web-based services is becoming a trend in application development. Many APIs for web-based application development provide a Representational State Transfer (REST) API. REST is a Web 2.0 standard [7]. REST describes an approach for a client/server architecture which provides a simple communication interface using XML and HTTP. In the REST specification, every resource is identified by a URI and the use of HTTP enables a software developer to communicate through simple GET, PUT and POST commands. This section provides summarised descriptions of the current specifications of the Twitter API and Facebook API. It should be noted that these APIs tend to evolve very quickly.

### **The Twitter API**

Twitter currently provides a streaming API and two discrete REST APIs [2, 31]. Through the streaming API, called the Firehose [2, 22, 31], users can obtain real-time access to tweets in a sampled and filtered form. The API is HTTP based and it supports the use of GET, POST and DELETE requests for data access. In Twitter terminology, individual messages describe the ‘status’ of a user. Using the streaming API, users can access subsets of public status descriptions in almost real-time including replies and mentions created by public accounts. The streaming API uses basic HTTP for authentication and requires a valid Twitter account. Data can be retrieved in XML or JSON format. The JSON format is very simple and can be parsed very easily because every line terminated by a carriage return contains an

object. The Twitter API allows the integration of Twitter with other web services and applications [20].

### **The Facebook API**

The Facebook API [15] was launched in 2007 and currently consists of five components as follows: (1) an HTML-based markup language called the Facebook Markup Language (FBML), (2) a REST API, (3) SQL-style query language for interacting with Facebook called the Facebook Query language (FQL), (4) a scripting language called Facebook JavaScript, and (5) a set of client programming libraries. The Facebook API enables developers to create external applications to empower Facebook users to interact with one another in new and exciting ways that are invented by the developers. In order to access Facebook data or develop a Facebook application, a developer needs a Facebook account. Obtaining Facebook data for offline analysis is a fairly straight forward matter. One uses the REST API and FQL to obtain Facebook data. Developing a Facebook application however requires a fairly high level of programming expertise. However, Facebook provides many learning aids to help developers to master the use of the API [15]. Since Facebook only provides methods for accessing data and displaying some information to the application user, it is the developer's responsibility to host the application. There are web sites that host Facebook applications for free, although there may be a waiting period after the developer applies for the free hosting [15]. After the application has been developed, hosted at a website and registered (by following the steps to create an application on Facebook and agreeing to the terms of service), Facebook will provide the application to other Facebook users when requested [15].

## ***6.3 Analytics for Social Network Data***

Some of the methods of knowledge discovery, data mining that have been applied to social network data are graph mining and sentiment analysis and clustering [2, 34]. This section provides a brief discussion of sentiment analysis, time series analysis and graph mining.

### **Sentiment Analysis**

Sentiment analysis of text messages may be defined as a classification problem where the task is to classify the messages into three categories depending on whether they convey positive, negative or neutral feelings [2, 30]. From a machine learning and data mining perspective, sentiment analysis involves the creation of a classification model which can be used to assign class labels (positive, negative, neutral) to messages. Commonly used algorithms for sentiment analysis are Naïve

Bayes, maximum entropy, support vector machines, and classification trees [2, 30, 34]. Most data mining and statistical software provide these algorithms. Text mining involves the analysis of the message text. The data mining methods that have been used for Twitter text mining include sentiment analysis through classification of tweets, clustering of tweets and trending topic detection [2]. Wakade et al. [34] have discussed the use of sentiment analysis for Twitter data. They have provided a list of activities necessary for sentiment analysis. These activities are given in Table 3.

Activity 1 in Table 3 (data collection) is achieved through the use of the Twitter API to obtain the data about a specific topic. This data may be stored in a document database. Activity 2 requires the use of specialised tools to conduct the data pre-processing activities. The WEKA software [16] provides tools for performing all the tasks listed under Activity 2 [2, 34]. Wakade et al. [34] have observed that two main challenges in the pre-processing of Twitter data are due to the usage of abbreviations (e.g. 'afaik' for 'as far as I know' and 'lol' for 'laugh out loud') and the usage of slang with different dialects such as netspeak and chatspeak. Additionally, it has been observed that in African countries, Twitter users tend to mix English words, French words with words from African languages such as Swahili, Zulu, and many others. One way to address these challenges is to compile additional lists of positive and negative words in African languages. This adds a local context component to the sentiment analysis activities of social network messages. Wakade et al. [34] have proposed the use of an additional pre-processing step to expand well-known abbreviations in tweets. Activity 3 (feature determination) and Activity 4 (sentiment labelling) may require the writing of specialised computer programs to conduct the tasks for the activities. Activity 5 (creation of classification model) can be performed using available data mining software e.g. WEKA [16] and Massive Online Analytics (MOA) [3], or statistical software, e.g. System R [18]. These three applications are FLOSS software and may be downloaded from the Internet.

Usage of the classification model can be conducted using the data stream mining approach. In this approach data (tweets) will be classified as they arrive [2]. The challenge is then to provide effective means of visualisation of classification results by human users. Several writers have observed that the effect of one tweet may be negligible but the effect of many tweets can be significant [19]. The MOA software [2, 3] can be employed for the classification of Twitter data using a classifier created with WEKA and MOA [2].

### **Time Series Analysis of Aggregate Sentiment**

A simple and effective method of providing the results of sentiment analysis classification results to human users in an organisation is through the computation of aggregates (e.g. daily aggregates) which can additionally be displayed graphically, e.g. using line plots or bar charts. Useful aggregates would be the counts or percentages for positive and negative sentiment tweets, or the ratio of positive to negative tweets on a given day. Another even more advanced practice is to use time series analysis. O'Connor et al. [24] have argued that it may be the case that on a day-to-day basis, the aggregate sentiment values may be highly volatile so that it

**Table 3** Activities for the creation of a classification model for sentiment analysis of tweets

Activity	Description
1. Data collection	Collect related tweets using a query containing words or phrases denoting the subject (topic) of interest
2. Data pre-processing	<p>Preprocess the tweet data as follows:</p> <ol style="list-style-type: none"> <li>(1) Remove the stop words. These are words that do not convey any positive or negative meaning. Examples are: is, was, been</li> <li>(2) Replace the URLs with a specific tag. It is common to find URLs in tweets as people share interesting links with friends</li> <li>(3) Remove symbols except those that make up emoticons</li> <li>(4) Stem the words. This is the process of reducing a word to its root form, e.g. the words <i>reader</i>, <i>reading</i> both reduce to the root word <i>read</i></li> <li>(5) Use the resulting data to create the training dataset from the classification model and a test dataset</li> </ol>
3. Feature determination	<p>Determine the features (predictor variables) for the classification model</p> <ol style="list-style-type: none"> <li>(1) Obtain a list of positive words (e.g. from <a href="http://www.inspiration.co.uk/positive.htm">http://www.inspiration.co.uk/positive.htm</a>) and a list of negative words (e.g. from <a href="http://eqi.org/fw_neg.htm">http://eqi.org/fw_neg.htm</a>)</li> <li>(2) Decide on specific symbols to represent emoticons</li> <li>(3) Combine the positive and negative words as well as the emoticon symbols into the features list</li> <li>(4) Reduce the length of the feature list. Some words in the features list may appear with a very low frequency in the tweets selected for the training data. Use a cut-off value (e.g. frequency &gt; 2) to select the words that appear in the final features list</li> </ol>
4. Sentiment labelling	<p>Each tweet in the training data needs to be labelled as positive, negative or neutral. The procedure used by Wakade et al. [34] is as follows:</p> <ol style="list-style-type: none"> <li>(1) For each tweet, count the number of positive words (<i>poscount</i>) and the number of negative words (<i>negcount</i>)</li> <li>(2) For each tweet determine whether there is an positive emoticon (<i>hasSmiley</i>) or a negative emoticon (<i>hasSad</i>) or no emoticon (<i>hasNone</i>)</li> <li>(3) Use the following logic to label each tweet:            Label a tweet as positive if:  <math>(poscount &gt; negcount) \text{ and } (hasSmiley \text{ or } hasNone)</math>            or <math>(poscount = negcount) \text{ and } (hasSmiley)</math>            Label a tweet as negative if:  <math>(poscount &lt; negCount) \text{ and } (hasSad \text{ or } hasNone)</math> or  <math>(poscount = negCount) \text{ and } (hasSad)</math>            otherwise label the tweet as neutral</li> </ol>
5. Creation of classification model	Using a classification algorithm, e.g. Naive Bayes, maximum entropy, support vector machines, or classification tree, and the features list, create a predictive classification model and test the model performance using the test dataset



is difficult to determine the general trend for the sentiment measures. They have suggested that the computation of the moving average for the sentiment time series data provides a smoothed measure which makes it easier to observe any emerging trends. The moving average for a time series at time  $t$  is computed as

$$MA_t = \frac{1}{k} (x_{t-k+1} + x_{t-k+2} + \dots + x_t)$$

where  $k$  is the number of past time periods (days in this case) and  $x_t$  is the value of the aggregate measure for the sentiment at time  $t$ . O'Connor et al. [24] have suggested the use of the ratio (positive to negative) as the aggregate measure for the moving average computations. When the values of the moving average are used for a line plot, the resulting plot is more smooth and more informative than when un-smoothed values are used. O'Connor et al. [24] have argued that smoothing is a critical issue as it causes a sentiment measure to respond more slowly to recent changes, and forces consistent behaviour to appear over long periods of time. It should be noted that too much smoothing (use of very large values of  $k$ ) makes it impossible to see fine-grained changes to the aggregate sentiment.

## Graph Mining

Graph mining is based on the analysis of links between social media users. Data for graph mining may be stored in a graph database (e.g. Neo4j). Graph mining of Twitter data involves the analysis of links between the messages [2]. Bifet and Frank [2] have reported that Twitter graph mining has been used to investigate interesting problems such as measuring user influence and the dynamics of popularity [6], community discovery and community formation in social networks [21, 27], and social information diffusion [8]. These types of analysis can benefit organisations which have a social media presence, to better understand the characteristics of their followers, and to possibly target their most influential followers for purposes of enhancing their online word-of-mouth advertising.

For Twitter data, three measures of user influence that have been reported in the literature are: indegree, retweets and mentions. Indegree influence is the number of followers of a user and directly indicates the size of the audience of that user. Retweet influence is measured through the number of retweets containing a user's name. This measure indicates the ability of a user to generate content which has pass-along value. Mention influence is measured through the number of mentions containing a user's name. This measure indicates the ability of a user to engage other users in a conversation [6]. For Facebook data, the number of fans and the number of posts and comments have been reported as useful measures of influence [11].

## **7 Discussion and Recommendations**

Social media adoption and usage, social networks and data analysis (online and offline) have been discussed in this chapter. This section provides a discussion of the benefits and challenges of using social media, and recommendations on how the challenges can be addressed.

### ***7.1 Benefits That Organisations Can Realise from Social Media Usage***

Public sector organisations and NGOs can benefit from the use for social media due to the low cost solutions for providing information to the public, engaging with and obtaining information from the public e.g. through polls and surveys, maintaining a presence on social networks and analysis of social networks data using free online tools. Big business and small business organisations can benefit from the use for social media due to the low cost solutions for marketing, branding and analysis of social media data using free online tools. Public sector organisations and big businesses can further benefit from the use for social media by conducting more sophisticated analysis offline using statistical and data mining tools. Every government has (or should have) a statistician general who heads a department of statisticians. Every big business organisation has a marketing department and public relations department that can dedicate resources to social media usage. However, there are still many challenges that need to be addressed before the benefits discussed above can become a reality for most African organisations. Some of these challenges are discussed below.

### ***7.2 Usage of Social Network Data Analytics in African Organisations***

It was stated in Section 4 that many organisations in Africa are actively using social media to engage with the public. FuseWare and World Wide Worx [11] have reported that in South Africa big businesses which are regarded as major brands rely heavily on social media agencies for social media content creation and social media monitoring. However, in 2014, 53% of these businesses plan to build up their social networking skills by investing in training for their marketing and public relations teams. FuseWare and World Wide Worx [11] have further reported that the measurement of social media effectiveness by big businesses remains relatively unsophisticated. For Twitter, 83% of these businesses measure effectiveness by the number of followers while only 48% conduct sentiment analysis. For Facebook, 87% of these businesses

measure the number of fans, 79% measure the number of posts and comments, and only 54% are assessing the tone of these posts and comments through sentiment analysis.

### ***7.3 Challenges in Social Media Usage***

One major challenge that has been highlighted by researchers is the shortage of expertise in effective social media usage in terms of designing content and conducting social media data analysis to assess the effectiveness of social media usage. This is especially true for African countries and for small businesses. For South Africa, FuseWare and World Wide Worx [11] have reported that, even though 91% of big businesses agree that social media has the potential for building a business, only 19% of these businesses believe that they are getting as much value from social media as they could. This is a strong indication that most organisations (small and large) are still learning how to effectively use social media in order to obtain value from this media. A second challenge that has been highlighted by researchers is the need, for organisations with a social media presence, to continuously monitor the social media services in order to respond to the public in a timely manner. FuseWare and World Wide Worx [11] have reported that, for the top-brand big businesses in South Africa, the average response time for addressing customer issues on Twitter is 271 minutes (4.5 hours). These authors have commented that: ‘Taking more than 4 hours to respond to a customer in such an immediate environment shows a gap in social media that needs to be closed’. One obvious conclusion that can be drawn from the foregoing discussion is that, if the big African businesses are still struggling to monitor and respond to social media communications in a timely manner, then small businesses and public sector organisations will also struggle with this aspect of social media usage.

Sections 5 and 6 provided a discussion of methods for obtaining social media data, storing this data in special databases, e.g. NoSQL databases, and conducting offline analysis of this data. At the present time, many organisations use relational (SQL) database, and these organisations possess IT expertise to conduct analysis of data in these databases. Technical expertise in the use of NoSQL databases is still very scarce in most African countries. This is partly due to the fact that university curricula on database systems largely concentrate on the relational database. Expertise in the use of social media APIs, e.g. the Twitter API and Facebook API is also very scarce. Additionally, the functionality of these APIs is changed very frequently by the social media services. In order to build capacity and expertise in conducting the types of social media data analysis discussed in Section 6, it will be necessary for universities and organisations to create educational programs that address these application development skills.

#### ***7.4 Research Perspective on Social Media Adoption and Data Analysis***

Academic literature on social media adoption and effective usage is hard to come by. It is a worthwhile challenge for African researchers in Computer Science, Information Systems and Statistics to engage in research which has the potential to create a knowledge base on effective social media usage in organisations. Effective sentiment analysis requires the use of sophisticated text mining methods which require (1) lists of positive and negative words (2) pre-processing of text to produce stemmed words, and (3) processing ability to determine the language for words that appear in mixed language (African and European) communications. It is a worthwhile challenge for African researchers in ICT-related fields to engage in inter-disciplinary research with academics in African languages (linguistics) in order to design and implement methods and tools for effectively processing African language text for purposes of text mining. The domain of Web 2.0 computing and social media is highly volatile. It is important for researchers to keep abreast of developments in this domain so that relevant research for economic development can continuously drive organisational understanding, increased adoption and effective usage of social media.

### **8 Conclusions**

This chapter has presented a discussion of the widely used social media and their adoption on the African continent, organisations on the African continent that have adopted these social media, how organisations that use social media can benefit from such usage, and the challenges posed by the adoption and usage of social media. Social media like Twitter, Facebook, LinkedIn and YouTube are widely used by many people all over the world, including African countries. Additionally, home-grown social media services such as Mxit are widely used in African countries. The major benefits of social media adoption by individuals, public sector and private sector organisations are the low cost solutions provided for exchanging information and analysing the effect of social media usage. Two major challenges for organisations have been identified in this chapter. The first challenge is the shortage of expertise in performing the tasks required for effective usage of social media. The required tasks include content creation and data analysis. The second challenge is the shortage of resources for continuous monitoring of social media communications from the public and provision of fast responses to these communications. In order to address these challenges, it will be necessary for African organisations and educational institutions to conduct capacity building activities that can lead to a reduction in the shortage of the required expertise. It will also be useful for African researchers to engage in research activities in support of effective and widespread usage of social media in government, non-government and business organisations.

## References

1. P. Anderson, What is Web 2.0? Ideas, technologies and implications for education. JISC Technology and Standards Watch (2007)
2. A. Bifet, E. Frank, Sentiment discovery in Twitter streaming data (2010)
3. A. Bifet, G. Holmes, R. Kirkby, B. Pfahringer, MOA: massive online analysis. J. Mach. Learn. Res. **11**, 1601–1604 (2010). <http://moa.cs.waikato.ac.nz>
4. Z. Bodnar, Generating small business customers with social media marketing: small business case studies (2014), Hubspot <http://www.HubSpot.com>. Cited 15 Mar 2014
5. C. Burke, Twitter search tools (2011), <http://charleneburke.com/2011/02/12-twitter-search-tools/>. Cited 28 Apr 2014
6. M. Cha, H. Haddadi, F. Benevenuto, K.P. Gummadi, Measuring user influence in Twitter: the million follower fallacy, in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, 2010, pp. 10–17
7. D. Dasgupta, R. Dasgupta, Social networks using Web 2.0. IBM DeveloperWorks (2009), <http://www.ibm.com/deveoperWorks/>
8. M. De Choudhury, Y.-R. Lin, H. Sundaram, K.S. Candan, L. Xie, A. Kelliher, How does the data sampling strategy impact the discovery of information diffusion in social media? in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, 2010, pp. 34–41
9. Deloitte and Frontier Advisory, Social Media in Africa, in *African Frontiers Forum* (Deloitte & Touche, Johannesburg, 2012), [http://www.deloitte.com/assets/Dcom-SouthAfrica/Local%20Assets/Documents/TMT\\_Social\\_Media.pdf..](http://www.deloitte.com/assets/Dcom-SouthAfrica/Local%20Assets/Documents/TMT_Social_Media.pdf..) Cited 28 Apr 2014
10. Free Software Foundation <http://www.fsf.org>. Cited 15 July 2014
11. Fuseware and World Wide Worx, South African Media Landscape 2014 – Executive Summary (2014), <http://www.worldwideworx.com/wp-content/uploads/2013/10/Exec-Summary-Social-Media-2014.pdf>. Cited 28 Apr 2014
12. Gartner, Gartner IT glossary (2013), <http://www.gartner.com/it-glossary/>. Cited 28 Apr 2014
13. GovernmentZA, Facebook page for the South African Government (2014a), <https://www.facebook.com/GovernmentZA>. Cited Mar 2014
14. GovernmentZA, Twitter profile for the South African Government (2014b), <https://mobile.twitter.com/GovernmentZA>. Cited 28 Apr 2014
15. W. Graham, *Facebook API Developers Guide* (Springer, New York, 2008). <http://www.apress.com>
16. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutenmann, I.H. Witten, The WEKA data mining software: an update. SIGKDD Explor. Newsletter **11**(1), 10–18 (2009)
17. HealthZA, Facebook page for the South African National Department of Health (2014), <https://www.facebook.com/HealthZA>. Cited 28 Apr 2014
18. K. Hornik, The R FAQ (2014), <http://CRAN.R-project.org/doc/FAQ/R-FAQ.html>. Cited 28 Apr 2014
19. IBM Global Business Services, Analytics: the real-world use of big data – how innovative enterprises extract value from uncertain data, IBM Institute for Business value. (2012)
20. B.J. Jansen, M. Zhang, K. Sobel, A. Chowdry, Micro-blogging as online word of mouth branding, in *Proceedings of the 27th International Conference Extended Abstracts on Human factors in Computing Systems*, Boston, Apr 2009, pp. 3856–3864
21. A. Java, X. Song, T. Finin, B. Tseng, Why we twitter: understanding microblogging usage and communities. in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, 2007, pp. 56–65
22. J. Kalucki. Twitter streaming API. <http://apiwiki.twitter.com/Streaming-APIDocumentation>. Cited 15 July 2014
23. A.B.M. Moniruzzaman, S.A. Hossain, NoSQL database: new era of databases for big data analytics – classification, characteristics and comparison. Int. J. Database Theor. Appl. **6**(4), 2013 (2013)

24. B. O'Connor, R. Balasubramanian, B.R. Routledge, N.A. Smith, From tweets to polls: linking text sentiment to public opinion time series, in *Proceedings of the AAAI Conference on Weblogs and Social Media*, Washington, May 2010, pp. 122–129
25. Z. Papacharissi, The virtual geographies of social networks: a comparative analysis of Facebook, LinkedIn and ASmallWorld. *New Media Soc.* **11**(1/2), 199–220 (2009)
26. PresidencyZA, Facebook page for the Presidency of the Republic of South African (2014), <https://www.facebook.com/PresidencyZA>. Cited Mar 2014
27. D.M. Romero, J. Kleinberg, The directed closure process in hybrid social information networks, with analysis of link formation on Twitter, in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, 2010, pp. 138–145
28. A. Sheombar, Social media use by Dutch development organisations: perceptions, applications and barriers. in *CDI Seminar*, Nov 2013
29. M.A. Stelzner, *2013 Social Media Marketing Report: How Marketers are Using Social Media to Grow Their Business*. Social Media Examiner. SocialmediaExaminer.com. (2013)
30. S.L. Ting, W.H. Ip, H.C. Tsang, Is Naive Bayes a good classifier for document classification? *Int. J. Software Eng. Appl.* **5**(3), 37–46 (2011)
31. Twitter, Twitter API. (2010), <http://apiwiki.twitter.com/>
32. Twitter, Twitter website: <http://www.twitter.com> (2014)
33. VentureBeat, Top 10 social media analytics tools: the VentureBeat index (2013), <http://venturebeat.com/2013/12/20/top-10-social-media-analytics-tools-the-venturebeat-index/>
34. S. Wakade, C. Shekar, K.J. Liszka, C.-C. Chan, Text mining for sentiment analysis of Twitter data, in *International Conference on Information and Knowledge Engineering (IKE'12)*, 2012, pp. 109–114
35. J. Yarow. Twitter finally reveals all its secret stats. BusinessInsider Weblog Article (2010), <http://www.businessinsider.com/twitter-stats-2010-4/>
36. YouTube Statistics (2014), <http://www.youtube.com/yt/press/statistics.html>. Cited 28 Apr 2014