International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 8 ǁ 2021 ǁ PP. 47-52

Identification of QR Code Perspective on enhancement of Text Mining Approaches

Dr. Rajeev Tripathi
Assistant Professor
SRMGPC
Lucknow, India

Dr. Santosh Kumar Dwivedi*
Associate Professor
SRMGPC
Lucknow, India

Abstract
Information sharing over a communication network has never been easier because to recent advances in communication technologies. Information is now processed electronically and disseminated over public networks. Because such networks are insecure, sensitive data must be safeguarded in some way. The study of mechanisms that allow us to achieve this is known as cryptography. Various cryptographic algorithms and firewalls are used to protect information from a variety of computer and network assaults. However, no single method can guarantee total safety. The information encoded within the QR code may be easily accessed using a decoding method that combines an image-capturing device with at least 300,000 pixels with the use of decoding software in the recognition of the QR code. Since the captured images can be rotated to the accurate angle prior decoding, the decoding is not restricted by the angle of the image. Digital video cameras are capable of observing QR code pictures. Text, URL links, and other forms of data can be stored in QR codes. They can be employed as a data carrier, allowing users to access technologies that are more comfortable to use.
Keywords: *QR Code, Text Mining, Data Mining*

Date of Submission: 16-07-2021 Date of acceptance: 01-08-2021

I. Introduction

A QR code is a type of digital barcode that encodes data as a series of pixels in a square-shaped grid and can be read simply by a digital device. QR codes are commonly used to track information about products in a supply chain, and they are regularly employed in marketing and advertising efforts because many smartphones come with built-in QR scanners. They've also been important in tracing coronavirus exposure and slowing the virus's spread in recent years.

QR codes are capable of storing a large amount of data. QR codes have been around for over fifteen years, thanks to a Japanese business. With the introduction of smart and Web-enabled mobile devices, we've seen a steady increase in the number of fascinating commercial QR code applications. The original QR code system was created in 1994 by Denso Wave, a Toyota subsidiary in Japan. During the manufacturing process, they needed a more precise technique to track cars and parts. They accomplished this by creating a barcode that could encode kanji, kana, and alphanumeric characters.

The QR code's creators sought to make it simple to scan so that operators didn't lose time trying to acquire it at the appropriate angle. They also wanted it to have a unique design that would make it easier to recognise. As a result, they chose the classic square design, which is still in use today.

Denso Wave released their QR code to the public and stated that they will not be using their patent rights. This means that QR codes could be created and used by anyone. However, regardless of how much information they include, when scanned, the QR code should allow the user to access information immediately - hence the name.

1.1 QR Code

Reaction Time: Despite many known applications for identifying 2D symbols, QR barcode identification in non-arbitrary environments remains a difficult task. The biggest drawback of modern QR code detection software is their poor performance when dealing with rotated and distorted single or multiple symbols in photos with varying lighting and noise.

Standard barcodes can only be read in one direction: from left to right. That implies they can only hold a limited amount of data, which is often in alphanumeric format. A QR code, on the other hand, may be read in two directions: top to bottom and right to left. This allows it to store a lot more information. Website URLs, phone numbers, and up to 4,000 characters of text can all be contained in a QR code.

QR codes exploded in popularity in South Korea and Japan a few years ago, but are only now beginning to catch on in the United States. While several modest independent studies on the efficiency of QR codes have been conducted in the United States, it wasn't until recently that a big study was conducted that provided important insight into the genuine usefulness of QR codes on various mediums.

Data mining techniques can be utilised in a healthcare application to detect unknown diseases, causes of diseases, and medical treatment procedures. It also aids medical researchers in building effective healthcare policies, drug recommendation systems, and individual health profiles. [1].

The barcode is a machine-readable visual representation of data about the object to which it is attached. Barcodes, also known as linear or one-dimensional codes, used to encode data by altering the lengths and spacings of parallel lines. Later, they evolved into rectangles, dots, hexagons, and other two-dimensional geometric designs.

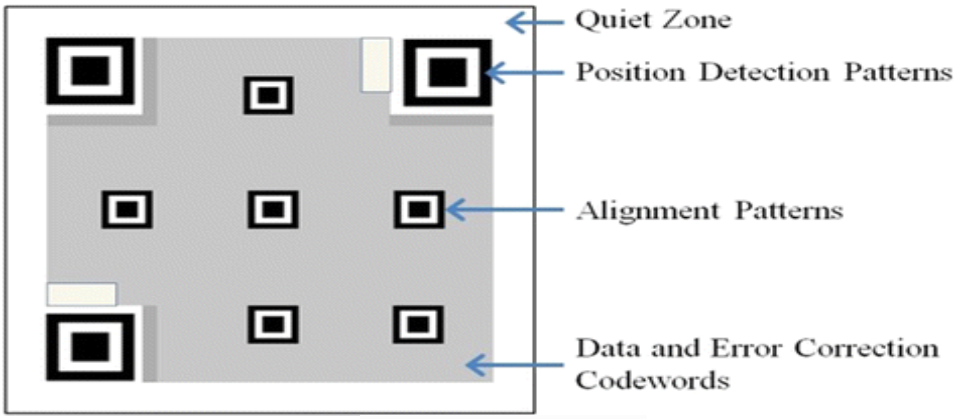


Fig. The modules of a QR code

The trademark for the form of matrix barcode invented by the Japanese business Denso Wave is QR code, which stands for Quick Response Code. Large capacity data encoding, filth and damage resistance, high speed scanning, tiny print out size, 360 degree reading, and structural freedom of application are just a few of the benefits of QR codes. The fast advancement of digital technology, image processing tools, and the web-based communication revolution has made the reproduction of digitally manufactured artifacts simple and within reach. Image processing is the process of converting a photograph into digital form and performing a few operations on it, all with the goal of obtaining a better photograph or extracting some useful information from it. Some of the outcomes were astonishing…

- An estimated 14 million mobile users in the U.S., representing 6.2% of all mobile users in the U.S., have scanned a QR code on their mobile devices.
- Almost half of all QR code users have scanned QR codes in print magazines or newspapers. QR codes featured on product packaging and websites were also scanned by over 25% of QR code users.
- Despite the heavy use of QR codes in print mediums such as fashion magazines, over 60% of the population who scanned QR codes were male.
- People scan QR codes primarily at home but are also likely to scan them in retail stores, grocery stores, and at work.
- Household incomes of people who scanned QR codes were predominately over $100,000 and the primary age demographic of code scanners were between the ages of 25 and 34.

Standard QR code is identifiable based on six components:

1. **Quiet Zone -** This is the empty white border around the outside of a QR code. Without this border, a QR reader will not be able to determine what is and is not contained within the QR code (due to interference from outside elements).
2. **Finder pattern -** QR codes usually contain three black squares in the bottom left, top left, and top right corners. These squares tell a QR reader that it is looking at a QR code and where the outside boundaries of the code lie.
3. **Alignment pattern** - This is another smaller square contained somewhere near the bottom right corner. It ensures that the QR code can be read, even if it is skewed or at an angle.

4. **Timing pattern -** This is an L-shaped line that runs between the three squares in the finder pattern. The timing pattern helps the reader identify individual squares within the whole code and makes it possible for a damaged QR code to be read.
5. **Version information -** This is a small field of information contained near the top–right finder pattern cell. This identifies which version of the QR code is being read (see "Types of QR code" below).
6. **Data cells -** The rest of the QR code communicates the actual information, i.e., the URL, phone number, or message it contains.

1.2 Types of QR code

Although QR codes can be used for a variety of applications, there are four widely acknowledged variations. The "input mode" governs how data can be saved and is determined by the version utilised. Numeric, alphanumeric, binary, or kanji can all be used. The version information section in the QR code communicates the mode type.

1. **Numeric mode -** This is for decimal digits 0 through 9. Numeric mode is the most effective storage mode, with up to 7,089 characters available.
2. **Alphanumeric mode -** This is for decimal digitals 0 through 9, plus uppercase letters A through Z, and symbols $, %, *, +, –, ., /, and : as well as a space. It allows up to 4,296 characters to be stored.
3. **Byte mode-** This is for characters from the ISO–8859–1 character set. It allows 2,953 characters to be stored.
4. **Kanji mode -** It is for double–byte characters from the Shift JIS character set and used to encode characters in Japanese. This is the original mode, first developed by Denso Wave. However, it has since become the least effective, with only 1,817 characters available for storage. A second kanji mode called **Extended Channel Interpretation (ECI) mode** can specify the kanji character set UTF–8. However, some newer QR code readers will not be able to read this character set.

There are two additional modes which are modifications of the other types:

- **Structured Append mode** - This encodes data across multiple QR codes, allowing up to 16 QR codes to be read simultaneously.
- **FNC1 mode -** This allows a QR code to function as a GS1 barcode.

II. Related Work

C. Baras and F. Cayre (2012) propose " the authentication problem of real-world goods on which 2D bar-codes (2D-BC) were printed and we take the opponent's point of view. A simple estimator of the 2D-BC based on copies averages is proposed, letting the opponent print a fake 2D-BC which aims at being declared as genuine by the system detector. Using an automated detection process based on a scan of the 2D-BC, a correlation score is computed and compared to a pre-determined threshold in order to decide whether the good is genuine or fake." [2]

T. V. Bui at all (2014), suggested that "Response Code (QR code) is widely used in daily life in recent years because it has high capacity encoding of data, damage resistance, fast decoding and other good characteristics. Since it is popular, people can use it to transmit secret information without inspection. Hiding secret information based on bit technique is so fragile to modification attack. If an attacker change any bit of hidden bits, it is impossible to recover the secret information." [3]

Falguni Patil, at all, 2015, "used the detection of QR codes, a type of 2D barcode, as described in the literature consists merely in the determination of the boundaries of the symbol region in images obtained with the specific intent of highlighting the symbol." [4]

III. Generate a QR Code in Python

qrtools contains a class QR (can be viewed in the source code), for which we must initially create an object. The object takes the following arguments

1. data
2. pixel_size
3. level
4. margin_size
5. data_type

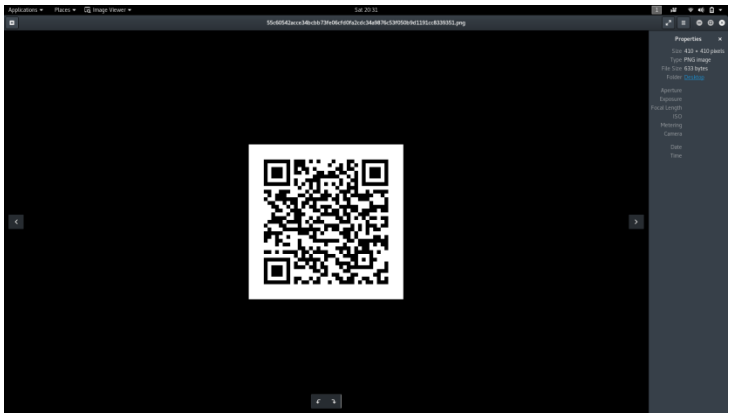
To create a QR code with default settings, we must simply identify the data while creating the object. Note that the data must be a unicode object if non-ASCII objects are going to be used.

```
# A Python program to generate QR code

from qrtools
```

```
import QR

# for creating the QR object
n_QR = QR(data = u"MYQRCode")

  n_QR = QR(data = u"MYQRCode", pixel_size = 10)

# encodes to a QR code
n_QR.encode()
```

If the program runs successfully, it returns a value of 0, and the QR code is stored in the tmp folder. To know the exact location, use the following command
print (newQR.filename)

Sample output
/tmp/qr-1496334996.385343/7489ebbcc2a00056ddaaaac190bce473e5c03696ea1bd8ed83
The pixel value of the QR code may also be changed by specifying the value during the creation of the QR object. The default size tends to be a little small for reading using scanners on smartphones, so a size of around 10 would be ideal for such purposes, for example:
The below QR code has pixel size = 10, and has been encoded with a URL



IV. Read a QR code

Scanning and reading a QR code is relatively simple. While creating the QR object, we must simply specify the path to the QR code as an argument.
Python program to Scan and Read a QR code

```
from qrtools import QR
nQR = QR(filename = "home/user/myfolder/myqr.png")
# decodes the QR code and returns True if successful
nQR.decode()
# prints the data
print (nQR.data)
```

V. Impacts QR code in Data Mining Activities

A number of data mining techniques are there like classification, clustering, advanced neural networks, prediction and regression models used for different data mining approaches in various areas[5]. One of the most important study fields in identifying relevant information from large databases is data mining. Web data, e-commerce, and other types of data are being collected and stored in large quantities. During recent years, there are major developments in the adoption of 2D Codes such as: The directive by International Air Transport Association (IATA) for airports worldwide to adopt 2D bar code for passenger boarding passes by 2010.The adoption of QR Code for patient identification by two leading hospitals in Singapore and all hospitals in Hong Kong. The use of 2D bar codes/micro codes for various applications is in the other sector. The use of QR code with mobile phones in Japan and Korea. Examples of such applications are: Large scale QR Codes on buildings to enable users to use mobile phone to scan the QR Code to retrieve information about the companies that are operating inside the buildings. The use of mobile phone to scan the QR Code on the packaging of fruits or vegetables to reclaim information about the name of the farm from which the fruits and vegetables are grown

and harvested; also the fertilizers and insecticide used. The QR Codes on the food packages when scanned will also enable consumers to download information on cooking recipes. QR Codes for location based services on maps in the Tokyo subway and central bus stations. Passengers can use their mobile phones to scan the QR Code to find out the arrival time of the next bus. Mobile phone and QR Code for payment of tickets for Trains and Airlines services. QR Code for TV program guides using mobile phone to view the program captured in QR Code.

5.1 K-Means

K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. K-means clustering aims to partition n observations into k clusters. In which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. Extracting information through interactive design of quires can achieve highly precise results in a short amount of time. Mush of time in this work was spent on pre-processing documents to allow the results to confirm specified format.[6]

The billing based system uses the functions of kmeans clustering algorithm. The billing based system uses k-means clustering in owner's database for storing the details of the customer who has purchased the product with details of the purchase. The owner can check whoever has purchased along with the billing part. Details of both direct paid customers and paypal way of money transacted customers will be displayed when the bill number is entered.

Other applications of k-means clustering are, in computer graphics, color quantization is the task of reducing the color palette of an image to a fixed number of colors k. In cluster analysis, the k-means algorithm can be used to partition the input data set into k partitions.

Algorithm: *k*-means. The *k*-means algorithm for partitioning, where each cluster's center is represented by the mean value of the objects in the cluster.

Input:
k: the number of clusters,
D: a data set containing n objects.
Output: A set of k clusters.
Method:
(1) Randomly choose k objects from D as the initial cluster centers;
(2) Repeat
(3) Re-assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
(4) Update the cluster means, i.e., calculate the mean value of the objects for each clusters; (5) until no change;

VI. Conclusion

Quick Response Code is another form of barcode. They can be used for business cards and other bits where people scan them into a system. QR Code is a two dimensional symbol. A real time capturing system for customer supplies using Quick Response (QR) code in Android smart phone. QR code verifies products by capturing it through the smart phone, then decodes and sends it to the server for authentication. The customer forwards the selected product list to the server and the response received from the server enables the consumer to decide based on the products authenticity. An interesting future study might involve simulating payment method at different gateway.

References

[1]. Koh, H. C., Tan, G.Data mining applications in healthcareJournal of Healthcare Information Management200519264722-s2.0-19844377332

[2]. C. Baras and F. Cayre. "2D bar-codes for authentication: A security approach". In Signal Processing Conference (EUSIPCO), proceedings of the 20th European, pages 1760-1766, 2012.

[3]. T. V. Bui, N. K. Vu, T. T. P. Nguyen, I. Echizen and T. D. Nguyen."Robust message hiding for QR Code". In Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2014 Tenth International Conference on, pages 520-523 IEEE, 2014

[4]. FalguniPatil, UtkarshaBhandari, and MadhuriKasar."QR Code Approach for Examination Process".International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169 Volume: 3 Issue: 2633-636, 2015.

[5]. R. Tripathi & S. Dwivedi "A Quick Review of Data Stream Mining Algorithms", IJIR, Vol-2, Issue-7, 2016.

[6]. R. Tripathi & S. Dwivedi, "Accurate Career Trends Extraction for Information Professionals using Agile Text Mining", IJARCSSE, Volume 5, Issue 12, December 2015

[7]. S. Rungraungsilp, M. Ketcham, V. Kosolvijak, and S. Vongpradhip, "Data hiding method for QR code based on watermark by compare DCT with DFT domain," in Proceedings of the 3rd international conference on computer and communication technologies, pp. 144–148, India, 2012.

[8]. Bughin, J., Livingston, J., & Marwaha, S. (2011). Seizing the potential of ‗big data'. The McKinsey Quarterly, 4, 103–109.Google Scholar

[9]. Carbone, P.L. (2000) Expanding the Meaning and Application of Data Mining. International Conference on Systems, Man and Cybernetics,3, 1872-1873.

[10]. M. and Heckerman, D., "An experimental comparison of several clustering and initialization method," Technical Report MSR TR-98-06, Microsoft

[11]. A Novel Method for text and Non-Text segmentation in document images. International Conference on Communication and signal processing, April 3-5, 2013 978-1-4673-4866-9/13 2013 IEEE.