

Creating Digital Archives of Ancient Scriptures using CAPTCHAs

Harkiran Kaur
Thapar University, Patiala
Deptt. of Comp. Sc. & Engg.
Thapar University, Patiala

Seema Bawa, PhD.
Thapar University, Patiala
Deptt. of Comp. Sc. & Engg.
Thapar University, Patiala

ABSTRACT

This paper proposes the use of CAPTCHAs for digitization of ancient scriptures available in the form of old handwritten as well as printed manuscripts. Two different approaches have been proposed in this paper, for digitization of these two forms of scriptures. The rare collection of these manuscripts will be lost, if not attended and preserved for future generations. Digitization is one such method to preserve our valuable heritage.

Keywords

CAPTCHA; digitization; scriptures; archives.

1. INTRODUCTION

“Completely Automated Public Turing test to tell Computers and Humans Apart” (CAPTCHA) is kind of Challenge - Response test used in computing, in order to ensure that the response is generated by a person. [1]

CAPTCHAs provide a problem, easy enough for humans to solve. It prevents automated software from filling out the form (hackers). CAPTCHAs are designed to keep away malicious computer "bots" and other automated programs, from using Web forms (to gather information about a server, initiate a cyber attack, or participate in phishing or spamming).

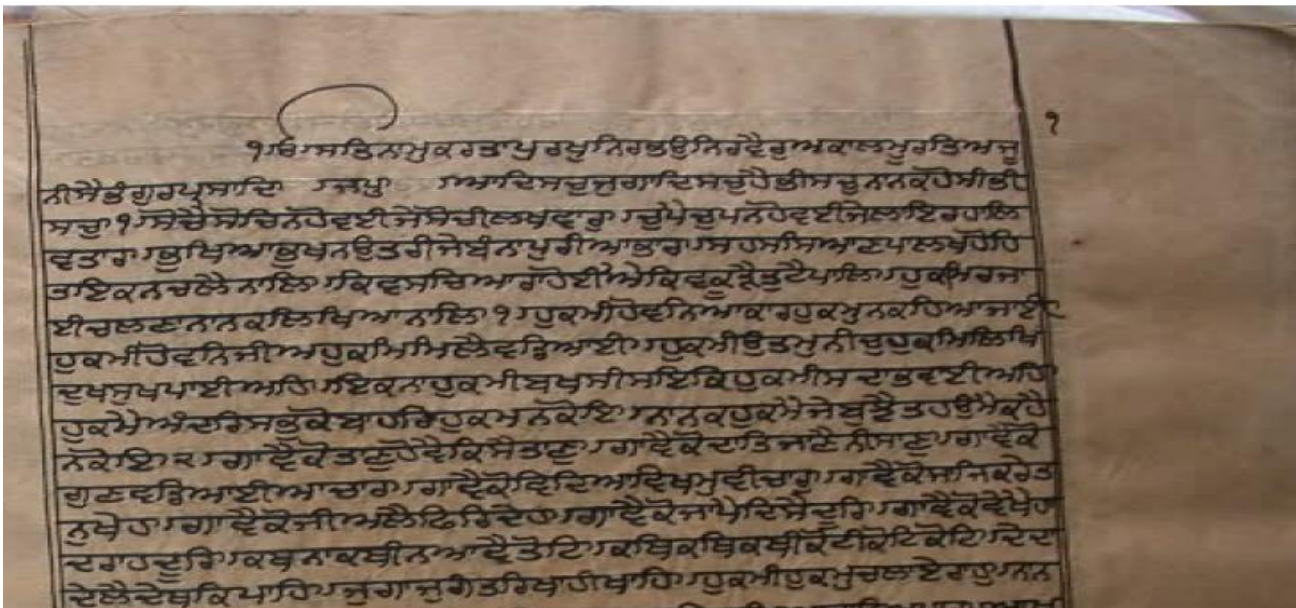


Fig 1: Handwritten text from Sikh Scripture Sri Guru Granth Sahib (SGGS)

CAPTCHAs are used for security on the website. Several companies like Yahoo!, Google, Microsoft, etc. offer free email services. CAPTCHAs prevent these email services from a specific type of attack - "bots" i.e. automated computer programs create a large number of email accounts, that is, protecting website registration.

Many of the bloggers suffer with the programs that submit bogus comments (comment spam). CAPTCHAs help in prevention of comment spam in Blogs.

Another scenario is of password systems, when the user enters wrong password again and again, a number of times. Instead of locking the user's account after certain number of unsuccessful logins; the better approach is to let the user solve a CAPTCHA after a certain number of unsuccessful logins. This will lead to prevention of dictionary attacks by the use of CAPTCHAs. Figure 1 represents an example of handwritten text from Sikh Scripture, Sri Guru Granth Sahib.

Figure 2 represents an example of printed text from Sikh Scripture Sri Guru Granth Sahib.

2. COMPARISON OF EXISITING USAGE of CAPTCHAs

CAPTCHAs are currently being used for implementing security on the websites. Usage of CAPTCHAs for security on a website includes:

Table 1: Comparison of existing and proposed usage of CAPTCHAs [1]

Existing Usage of CAPTCHAs	Proposed Usage of CAPTCHAs
a. To protect system vulnerabilities to e-mail spam. Example – e-mail services of Gmail, Yahoo mail, etc.	a. Digitization of scriptures.
b. To minimize automated posting to Blogs, wikis etc.	b. Storing the scriptures for future generations.
c. To confirm that the task is being performed by a human not an automated program.	c. Preserve ancient scriptures for forever.



Fig 2: Printed text from Sikh Scripture Sri Guru Granth Sahib (SGGS) [2]

3. USE of CAPTCHAs in CULTURAL COMPUTING

There is diversity in terms of culture in India. [4] Cultural Computing is important due to certain reasons - The young generation today, remains deprived of the knowledge of their cultures, due to various changes in their lifestyle, like the trend of nuclear families, parents working, changes in education system, etc.

Further, it is very difficult to understand any culture, by just reading books or listening to lectures related to it; as every culture has its roots in locality and nationality. [6]

Also, some of the cultural contents are not receiving much attention and they are getting deteriorated day by day. So, the need of the hour is to digitize and preserve cultural objects for our future generations to know about their cultures, to

understand these cultures in a simple and easy manner and to preserve them for forever. [3]

One of the main objectives of Cultural Computing is the preservation of cultural heritage contents, through digitization of these digital cultural objects. Cultural Computing is an umbrella under which all these objectives can be fulfilled. [5]

This paper proposes the use of CAPTCHAs for digitization of the old scriptures, related to Sikhism. And similar effort is needed for other cultures' ancient scriptures also. One of the methods, for digitization of the cultural objects (like ancient scriptures related to a culture) is the use of CAPTCHAs.

The CAPTCHAs used on a cultural portal for implementing the concepts of Cultural Computing, will contribute to the digitization of those ancient scriptures. The ancient scriptures related to Sikhism are available in both handwritten form and in printed form. Separate approaches are proposed for the use of CAPTCHAs for both the printed and the handwritten scriptures.

CAPTCHA is like a challenge-response test or challenge-response authentication. That is, one party presents a challenge and another party must provide the valid response (which will be authenticated). [8]

Example of challenge response test is password authentication. Here, challenge is asking for password and response is filling of correct password. A CAPTCHA challenges the user to type letters or digits (appearing in the form of a distorted image that appears on the screen, to the user) which the computer programs can't understand.

3.1 CAPTCHAs for processing printed ancient scriptures

Many of the scriptures of Sikhism are also available in printed form.

As the youth generally do not have interest in reading scriptures. And youth generally gathers much of the information required on any topic, by surfing internet.

It is the high time to make the youth aware about these scriptures, through digitization of these scriptures. That is, make them understand, through the websites they visit.

Member's Login [ਮੈਂਬਰ ਲਾਗ ਏਨ]

User Id: harkiran@sikh.org

Password:

Enter code (in lowercase letters): ਸਤਿ

Gurmukhi Keyboard

,	ੴ	ੲ	ੳ	ੴ	ੲ	ੳ	ੴ	ੲ	ੳ	ੴ
<- Backspace			ੴ	ਅ	ੲ	ਸ	ਹ	ਕ	ਖ	ਗ
ਘ	ਙ	ਚ	ਛ	ਜ	ਝ	ਵ	ਟ	ਠ	ਡ	ਢ
ਣ	ਤ	ਥ	ਦ	ਧ	ਨ	ਪ	ਫ	ਬ	ਭ	ਮ
ਯ	ਰ	ਲ	ਵ	ੜ	ੳ	ੳ	ੳ	ਆ	ਐ	ਐ
ਇ	ਈ	ੲ	ੲ	ੲ	ੲ	ੲ	ੲ	ੲ	ੲ	ੲ
ੳ	ੳ	ੳ	ੳ	ੳ	ੳ	-	=	.	,	:
?	;			ੴ			ੴ			

LOGIN

Fig 3(a): CAPTCHA generation (from printed SGGS Scripture) on Login

The CAPTCHAs created from the printed Sri Guru Granth Sahib are taken as an example (as shown in figure 3(a)) to explain the idea clearly.

These CAPTCHAs are used in the login system. After the login screen appears, user enters his user id and password and the code as shown in the dynamically generated image called CAPTCHA. The chances of error in entering printed CAPTCHA code are less as compared to handwritten CAPTCHAs. The user will have to enter the code as visible in the CAPTCHA image and click on 'Login' button.

Each CAPTCHA will be visible only once to every user of the cultural portal. The Login Process will be completed only after entering the CAPTCHA code, as shown in Figure 3(a).

One page of (this copy of) printed Sikh scripture, Sri Guru Granth Sahib, gets digitized in 154 executions of the code, used for the purpose. The processing of printed CAPTCHA is determined with the help of snapshots shown in Figure 3(a), Figure 3(b) and Figure 3(c).

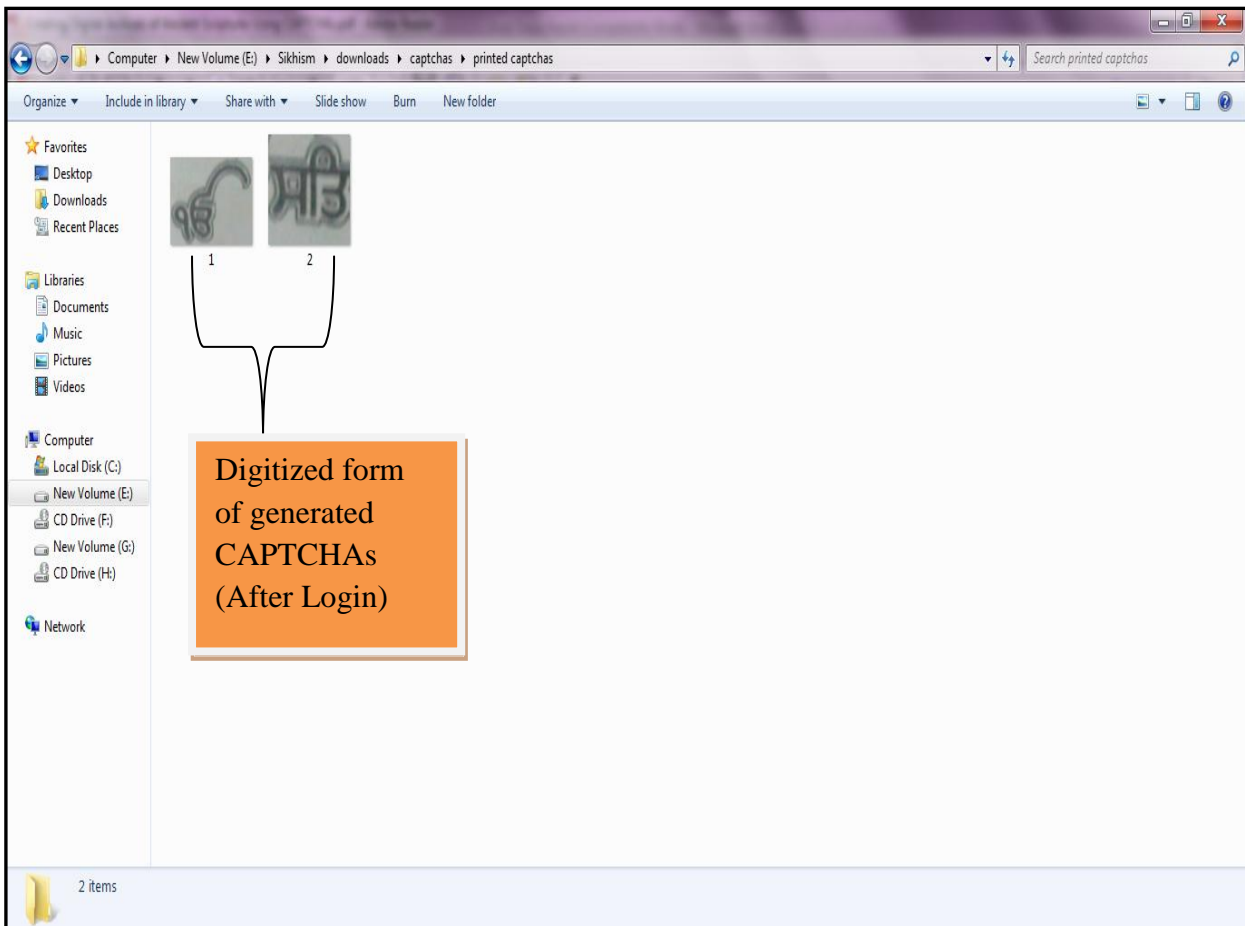


Figure 3(b): Digitized CAPTCHAs of Printed SGGS Scripture


Figure 3(a) shows dynamic generation of CAPTCHA in sequence (of words-page wise) of a printed scripture of Sri Guru Granth Sahib.

Figure 3(b) depicts the digitized form of the scripture after each dynamic generation of the CAPTCHA. After each CAPTCHA is processed, it gets stored in a particular folder as shown in Figure 3 (b). In this way, complete scripture can be digitized and preserved for future generations.

If the user enters the code wrong, then a message appears 'Code Entered incorrectly', as represented in Figure 3(c), the validation process of a printed CAPTCHA. And the user will again have to enter the code and press 'Login' button.

Figure 4 represents a flowchart showing the complete processing of printed CAPTCHA.

Member's Login [ਮੈਂਬਰ ਲਾਗ ਏਨ]

User Id	harkiran@sikh.org
Password	
	
Enter code (in lowercase letters)	ਨਾਮ

CAPTCHA entered incorrectly by the user

Gurmukhi Keyboard										
ੴ	ਕ	ਖ	ਗ	ਘ	ਙ	ਚ	ਛ	ਜ	ਝ	ਞ
<- Backspace	ੳ	ਅ	ੲ	ਸ	ਹ	ਕ	ਖ	ਗ		
ਘ	ਙ	ਚ	ਛ	ਜ	ਝ	ਞ	ਟ	ਠ	ਡ	ਢ
ਠ	ਡ	ਢ	ਪ	ਫ	ਬ	ਭ	ਮ			
ਯ	ਰ	ਲ	ਵ	ਸ਼	ੲ	ੳ	ਆ	ਐ	ਓ	
ੲ	ੳ	ੴ	ੵ	੶	੷	੸	੹	੺	੻	੼
?	:		=							

Code Entered Incorrectly!

[LOGIN](#)

Fig 3(c): Validation Process of Printed CAPTCHA

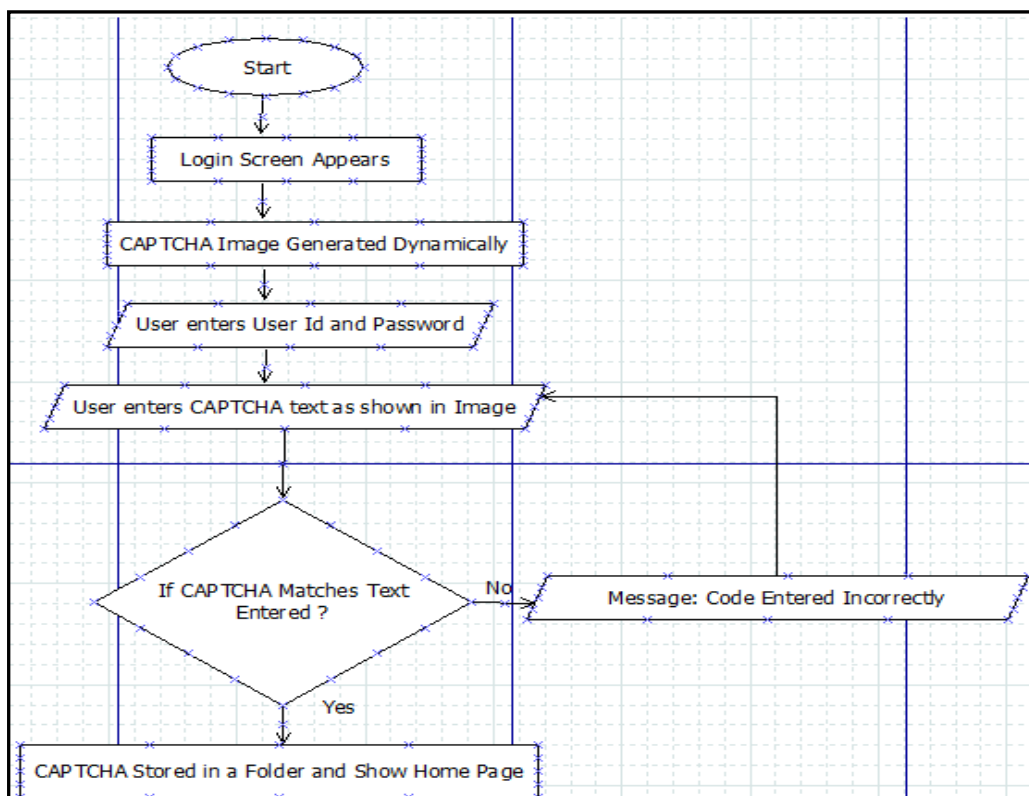


Fig 4: Flowchart representing processing of Printed CAPTCHA

3.2 CAPTCHAs for processing handwritten ancient scriptures

The handwritten scriptures of Sikhism are not receiving much attention and our young generation is not even aware about their whereabouts. So, in order to preserve these ancient scriptures, and make them available for our future generations; they need to be digitized. Otherwise, the future generations will remain deprived of the knowledge of these cultural objects.

For the digitization of these handwritten scriptures, one of the methods as proposed in this paper, is creating CAPTCHAs of these scriptures.

In this paper, CAPTCHAs of Sri Guru Granth Sahib are taken as an example, to explain the idea. CAPTCHAs are used in password systems. So, a number of users can be created by providing sign up facility on the cultural portal. Whenever, a new user is created or the existing user logs in, the user is required to enter the code shown in the form of a distorted image.

Each CAPTCHA of the scripture is visible only once to every user of the cultural portal. The Login Process is completed only after entering the CAPTCHA code. The code information (if entered wrong) by the user while solving the CAPTCHA is stored in the database. This code will be

verified by the administrator of the database. Administrator will observe if many of the users are experiencing errors while entering a particular CAPTCHA; then he will make changes in that very CAPTCHA to solve the visibility issue. The idea is shown in the snapshots of screens in the Figure 5(a), 5(b) and 5 (c).

After the user logs in (by entering the password), a screen will appear, showing the handwritten manuscript's words (CAPTCHA). Figure 5(a) depicts the dynamic generation of CAPTCHA from handwritten Sikh scripture SGGs. The logged in user will have to enter the CAPTCHA text in a textbox. If the word matches as entered by the user, login process becomes successful. In this way, a repository of the words of scripture can be stored and digitized.

This will add to the security feature of the cultural portal. As, the cultural portal will be prevented from an attack called bots; that is, the CAPTCHA will confirm that the user trying to login or create a new user account is a human, not an automated program.

One page of this copy (taken in consideration in this paper) of the handwritten form of the Sikh scripture, SGGs is digitized in 253 executions of the code for this purpose.

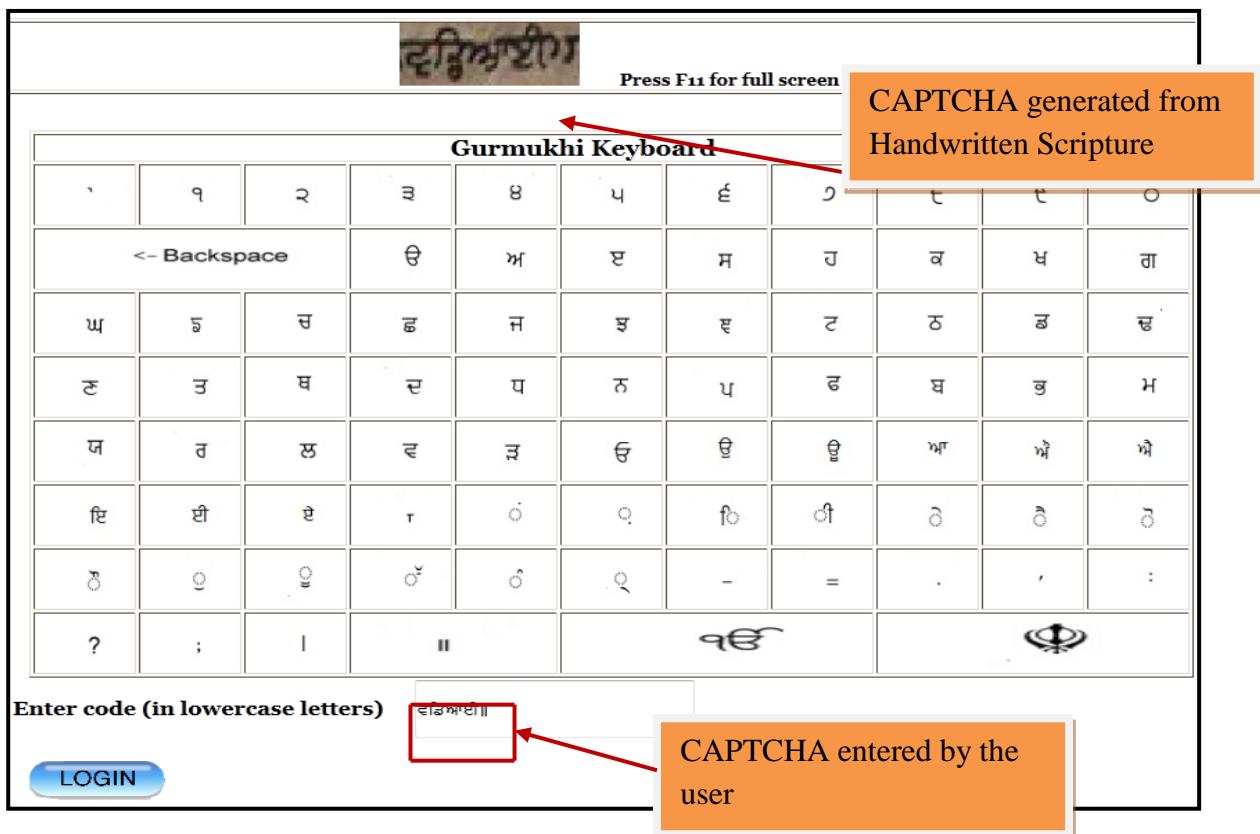


Fig 5(a): CAPTCHA generation from handwritten Sikh scripture SGGs

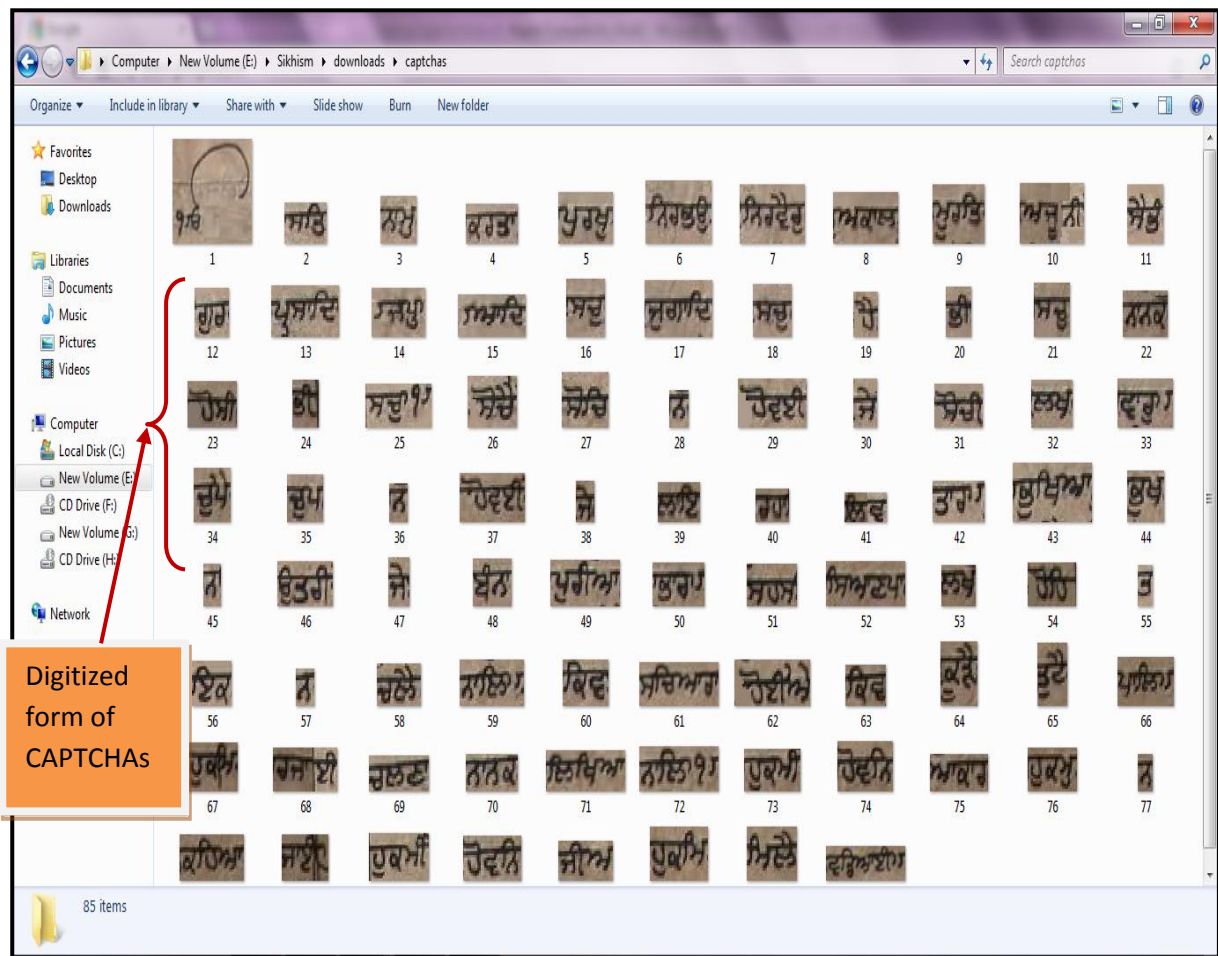


Fig 5(b): Digitized CAPTCHAs of Handwritten SGGS Scripture

Figure 5(b) depicts the digitized form of the scripture after each dynamic generation of the CAPTCHA. If the user enters the code wrong, then a message appears 'Code Entered incorrectly', as represented in Figure 5(c).

When user clicks on 'Go Back' button, then new CAPTCHA image will be generated on the screen, to be entered by the user. The user is given three chances, to enter the code correctly. If the user enters the code wrong third time, the control will be redirected to Login page.



Figure 5(c) – Validation Process of Handwritten CAPTCHA

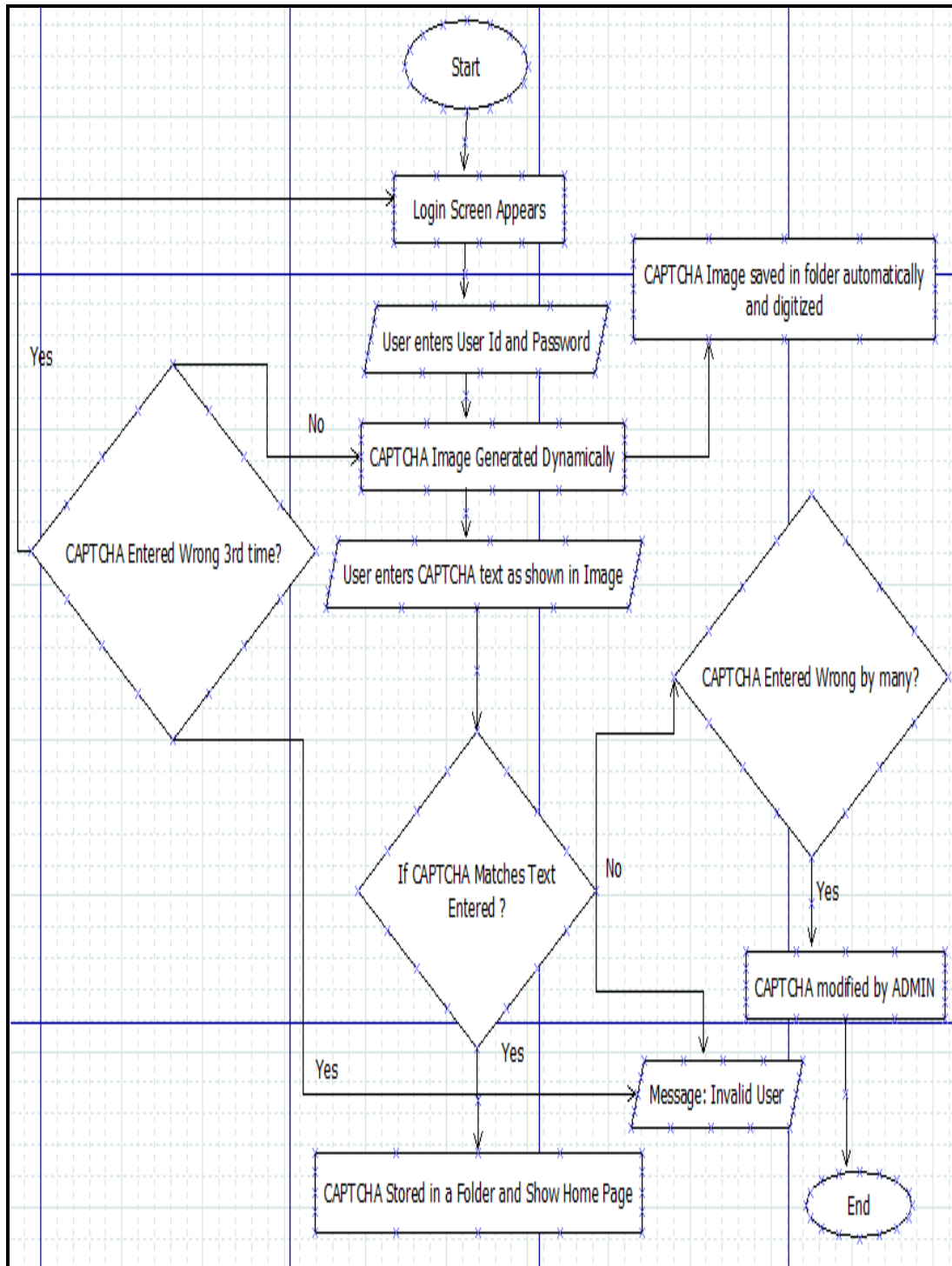


Fig 6: Flowchart representing processing of Handwritten CAPTCHA

4. CONCLUSION

This paper proposes use of CAPTCHAs. CAPTCHAs are generally used for security purpose on the websites. This paper has proposed the idea of using CAPTCHAs for security purpose (on the website) and simultaneous digitization of ancient scriptures.

The domain of this paper is scriptures related to Sikhism. These scriptures are available in both printed and handwritten form. Two different approaches for digitization have been presented in this paper for both these categories of the scriptures.

5. REFERENCES

- [1] 'CAPTCHA' at <http://en.wikipedia.org/wiki/CAPTCHA>
- [2] "Aad Sri Guru Granth Sahib" Published by B. Chattar Singh, Jiwan Singh.
- [3] Matthias Rauterberg, *From Personal to Cultural Computing: how to assess a cultural experience*, G. Kemper & P. von Hellberg (2006, eds.) uDayIV—Information nutzbar machen. Pabst Science Publ.
- [4] "The Cultural Heritage of India" Edited By – Haridas Bhattacharyya. Published by Swami Lokeswarananda, Secretary The Ramakrishna Mission Institute of Culture Gol Park, Calcutta, 1983.
- [5] Ryohei Nakatsu, Naoko Tosa, *A Computational Approach to Culture*, Cover Story at www.innovationmagazine.com
- [6] Kate Fernie, MLA, *The local dimension in the European Digital Library: from MINERVA and CALIMERA to MICHAEL*, Intelligent Access to Digital Heritage, Tallin, October 2007
- [7] Kashmir Singh, Gurdwara and the Aad Guru Granth Sahib: A Legal Perspective, Understanding Sikhism-The Research Journal
- [8] Rich Gossweiler, Maryam Kamvar, Shumeet Baluja, *What's Up CAPTCHA? A CAPTCHA Based On Image Orientation*, International World Wide Web Conference Committee (IW3C2) (2009)