

## A NEW ENCODING SCHEME OF TELUGU TEXT FOR INFORMATION HIDING

LAKSHMEESWARI G.\*, RAJYA LAKSHMI D. AND SRINIVAS Y.

GITAM Institute of Technology, GITAM University, Andhra Pradesh, India

\*Corresponding author. E-mail: \*lakshmeeswari@gitam.edu, rdavuluri@yahoo.com, ysrinivasit@rediffmail.com

Received: Received: August 10, 2011; Accepted: September 15, 2011

**Abstract-** Conveying information secretly and establishing a hidden relationship between the message and its counterpart has been of great interest since long time. This paper presents a new scheme for coding Telugu alphabet set, its diacritics and compounds which can be used in the Linguistic Steganography as well as cryptic transmission of Telugu text. This encoding scheme helps its users to have better support in implementing the hiding techniques. This scheme is also useful to other similar Indian Languages.

**Keywords:** Cryptic Algorithms, Steganography, Telugu text, diacritics, compounds, Information hiding.

### Introduction

Information hiding is the ability to prevent certain aspects of a component from being accessible to its clients. Since the rise of the Internet, one of the most important factors of information technology and communication has been the security of information [5]. Hiding information while communication takes place is an art. We may sometimes want to hide information in other information to make it unnoticeable during transmission. Some applications may use absolute invisibility of the secret information, but others require a larger message for a secret message to be hidden.

This paper deals with a novel scheme for encoding Telugu text which can be transmitted via network by encrypting it using any of the Crypting algorithms. We can transmit Telugu text in encoded form using the proposed scheme. We define a new encoding scheme suitable for Telugu language and use the generated code for Steganographic as well as Cryptic transformations.

Among the languages, after English, one of the most used South-Central language is Telugu. Hence, we considered Telugu language for this purpose as it is one of the most prominent languages used around the world. Telugu is the official language of the state of Andhra Pradesh. 73 million people use this language around the world, which include countries like Singapore, Phiji, U.S.A[4]. It occupies 15<sup>th</sup> position amongst the most widely used languages in the world and stands second in India. Telugu is a richly developed language and the biggest linguistic unit in India[4].

Organization of the remaining part of the paper is as follows: Section-2 contains description about the Historic evolution information of Telugu language, Section-3 briefs about the Telugu Literature, Section-4 presents a detailed description of the coding and conclusions are presented in Section-5.

### Historic Information of the Language

Telugu is a language derived from ancient Brahmi script. Brahmi based script is well known for its complex conjunct formations[8]. Telugu, Kannada and Tamil languages are called Dravidian languages[4]. These languages have complicated script. It was also referred to as 'Tenugu' in the past.

Telugu script is a combination of vowels, consonants, compounds and diacritics. Almost all Indian languages are built upon the concepts of vowels, consonants, diacritics and compounds and also have their origin from Sanskrit. Telugu comprises of 70% Sanskrit orientation[3]. Therefore the method that has been proposed below for Telugu language would well suit other Indian languages because Sanskrit is one of the base for all the Indian languages.

Telugu is one of the 22 official languages of India. Telugu also has official language status in the Yanam District of the Union Territory of Puducherry. Telugu was called the Italian of the East by Italian explorer Niccolò Da Conti[5].

Telugu script can reproduce the full range of Sanskrit phonetics without losing any of the text's originality. Telugu has made its letters expressive of all the sounds and hence it has to deal with significant borrowings from Sanskrit, Tamil and Hindustani[6].

### Literature survey

A *crypto system* transforms plain text messages (using a key) to un-intelligible text. *Cryptography* is the study of "secret writing" or cryptograms. Encryption is the process of converting plain text into cipher text. Decryption is the vice-versa of encryption i.e transforming the cipher text to original plain text[8].

Steganography's goal in general is to hide data well enough that unintended recipients do not suspect the

steganographic medium of containing hidden data[8]. Steganography is a technique of hiding information in digital media. In contrast to cryptography, it does not reveal others from knowing that the hidden information even exists. The goal of steganography is to avoid drawing suspicion to the existence of a hidden message. This approach of information hiding technique has become important in a number of application areas like Text, Digital audio, video, and pictures[ 7]. Many techniques are existing to hide English text into text of different languages like hiding English text in Bengali[2], Hindi[1], Urdu, Arabic text, etc. which use different techniques to achieve the same. The proposed system stores Telugu text in an English text or text of any other language or in a Digital Audio or Video. The main applicative part of this can be in the area of communication between the banker and its customer, between the corps for intelligence purposes. As Telugu is the local language of Andhra Pradesh, application can well suit the security necessities of the organizations requiring any authentication or hidden transmissions. Java Unicode occupies more number of bits when compared to the proposed scheme. As the requirements of Steganography states that the minimum the size of the cover text the best is the technique.

**Proposed coding scheme for Telugu text**

Telugu alphabet set has been consistently modified since its origin and presently it consists of 52 symbols - 16 vowels and 36 consonants. Sanskrit and Telugu alphabets are similar and exhibit one-one correspondence. The vowels are assigned values from 0 to 15 and the same range is considered for the diacritics also as diacritic is a combination of consonant with an vowel. The codes of compounds and constants are the same as the compound would always be a consonant. The code is a decimal number assigned uniquely for every alphabet of the language. Just varying the alphabets depending upon the language would make the coding scheme suitable for other languages.

**Vowels (Achulu)**

అ	ఆ	ఇ	ఈ
0	1	2	3
ఉ	ఊ	ఋ	ౠ
4	5	6	52
ఎ	ఏ	ఐ	ఒ
7	8	9	10
ఓ	ఔ	అం	అః
11	12	13	15

**Consonants (Hallulu)**

క	ఖ	గ	ఘ	ఙ
16	17	18	19	51
చ	ఛ	జ	ఝ	ఞ
20	21	22	23	50
ట	ఠ	డ	ఢ	ణ
24	25	26	27	28
త	థ	ద	ధ	న
29	30	31	32	33
ప	ఫ	బ	భ	మ
34	35	36	37	38
య	ర	ల	ళ	వ
39	40	41	42	43
శ	ష	స	హ	ఱ
44	45	46	47	48

**Diacritics( Gunintalu)**

√	ా	ి	ు	ి
0	1	2	3	4
ీ	ై	ౌ	ృ	ౠ
5	6	7	8	9
ౌ	ృ	ౠ	ౡ	ం
10	11	12	14	13
ః				
15				

**Compounds (Vattulu)**

కా	ఖా	గా	ఘా	ంగా
చా	ఛా	జా	ఝా	ఞా
టా	ఠా	డా	ఢా	ణా
తా	థా	దా	ధా	నా
పా	ఫా	బా	భా	మా
యా	రా	లా	ళా	వా
శా	షా	సా	హా	ఱా
కా	ఖా	గా	ఘా	ంగా
చా	ఛా	జా	ఝా	ఞా
టా	ఠా	డా	ఢా	ణా
తా	థా	దా	ధా	నా
పా	ఫా	బా	భా	మా
యా	రా	లా	ళా	వా
శా	షా	సా	హా	ఱా

Diacritics have same code as that of the vowels. Similarly, compounds have the same code as that of the consonants.

## Conclusion

In this paper a novel coding scheme has been proposed with the Telugu character set. The proposed codes can be extensively used for all data hiding purposes.

The coding scheme developed helps to minimize the bits required for hiding each character along with its diacritic and compound. We can hide a larger message with a minimal effort. It ensures multiple security levels for better secrecy of the content.

## References

- [1] Kalavathi Alla, Siva Rama Prasad R. (2009) *Sixth International Conference on Information Technology: New Generations*, 1577-1578.
- [2] Shri Changder S., Das S. and Shri Ghosh D. (2010) *2<sup>nd</sup> International Conference on Computer Technology and Development*, 501 – 505.
- [3] <http://www.balmitra.com/languagebook/telugu>.
- [4] An overview of the Telugu language is from [www.andhrabulletin.com/Telugu/Telugu\\_home.php](http://www.andhrabulletin.com/Telugu/Telugu_home.php)
- [5] [http://en.wikipedia.org/wiki/Telugu\\_languages](http://en.wikipedia.org/wiki/Telugu_languages)
- [6] History of the Telugu Language has been extracted from the reference <http://reference.findtarget.com/search/Telugu%20language/>
- [7] Artz D. (2001) *IEEE Internet Computing*, 75-80.
- [8] Chapman M. T. (1997) *"Hiding the hidden: A software system for concealing ciphertext as innocuous text"*, Master's thesis, University of Wisconsin-Milwaukee.



G. Lakshmeeswari, Asst. Prof., GIT, Gitam University, Visakhapatnam, AP, India. Her research area is Security. She has 12 years of experience and is now doing her Ph.D in Steganography.



Dr. D Rajya Lakshmi, Prof & HOD, GIT, GITAM University Visakhapatnam, AP, India. Her research interests include image processing, Data mining, Security and Computer Networks. She has 19 years of teaching and research



Dr. Y Srinivas, Professor, Dept. of Information Technology, GIT, GITAM University Visakhapatnam, AP, India. His research interests include image processing, Data mining, Security and Computer Networks. He has 19 years of teaching and research experience.