

## Research and Implementation of Ancient Yi Language Information Processing Platform in South Yunnan

Ying-Yu YANG<sup>1,2,3,a</sup>, Bing-Ze LI<sup>1,3,4,a,\*</sup>, Jia-Mei WANG<sup>1,3,4,a,\*</sup>, Guang-Ming LU<sup>1,3,4,a</sup>, Fei ZHAO<sup>1,3,4,a</sup>, Rui LIN<sup>1,3,4,a</sup> and Chang-Sen YUAN<sup>1,3,4,a</sup>

<sup>1</sup>School of Electrical Information Engineering, Yunnan MinZu University. Kunming, China, 650500

<sup>2</sup>University of Bath, The United Kingdom of Great Britain and Northern Ireland

<sup>3</sup>Yunnan Province colleges minority language information processing engineering research center, Kunming, China

<sup>4</sup>Provincial scientific research innovation team in Yunnan MinZu University

<sup>a</sup>1575763065@qq.com

\*Corresponding author

**Keywords:** Yunnan ancient Yi language; information processing platform; Liberal split encode; Digital Health Stroke encode.

**Abstract:** This paper attempts to analyze and research on Particularity of Yi character encoding information processing platform and communication system model in Windows operating system, develop the Yunnan ancient Yi language digital information processing platform. This platform consists of Barcode encoded input methods by standard Yi character Liberal split mode, handwritten Yi character Liberal split mode and standard Yi character Digital Health Stroke split mode, with the view of better inheritance of cultural characteristics of ethnic minorities in Yunnan and benefits of future generations.

### Introduction

Yi language belongs to the Sino-Tibetan Tibeto-Burman Yi branch. China's Yi language, points north, east, south, south, west, central six dialect, including five sub-dialects, namely Yunnan, Guizhou and eastern dialect is divided into sub-dialects and Northeast sub-dialects and sub-dialects Panxian; northern dialect is divided into northern and southern sub-dialect sub-dialects. Also points 25 dialect. Located in the northern dialect of large and small Liangshan region, eastern dialect of Sichuan, Yunnan, Guizhou, and distributed in the Northeast, the southern dialect distributed in Yunnan and Guangxi, Yunnan, and the remaining three are located in Yunnan dialect. With a population of about 400 million. Original one kind of Yi syllable word (it was considered ideographic), designed in 1956 in the form of the Latin alphabet phonetic Liangshan Yi program. 1975 Sichuan Liangshan Yi district of Sichuan developed "Yi specification pilot program", a total of 819 identified specification Yi words, in 1980 the State Council approved the implementation of the use of Sichuan Yi district. Yi language in foreign countries, mainly in Myanmar, Thailand and Vietnam.

Yunnan has a long history of ancient Yi culture, southern Yunnan region and preserve a richer cord books Yi, Yi culture has become an important part of the treasure trove. In order to rescue the national cultural treasure Yi soon, the paper studies the existing variety of input methods, such as "Sarbanes-Yi input method", the browser-based on line Yi input method, Yi Unicode input method input method, the YiYi Win input method input method, Chuxiong Yi Wen stroke input method, text input method Shilin, Vista system comes Yi input method, "American SIL Yi fonts and input method" and so on, based on the response to these input methods biggest flaw is unable to meet the needs of Yi ancient literature published. Font software developed with only 1165 words, cannot be covered by 2500 Yunnan Yi specification text, but cannot cover the tens of thousands of ancient Yi text, cannot meet the voluminous literature published Yi ancient problem. This paper developed an information theory-based, multi-layer, universal (Windows operating systems and publishing systems) are digitally encoded information processing platform, which are based on the

international standard Unicode encoding operating system designed to run on Windows 98 above, XP operating system, that in these operating systems are installed on the preparation of this article as long as the registration software, three input methods to install the software, southern Yunnan Yi True Type font software, you can achieve separately developed in this paper correspond to the three southern Yunnan Yi input information processing platform.

**The key theoretical issues Ancient Yi digitally encoded information processing platform needed to be addressed**

Plus minority language operating system in the Windows operating system, not only to solve technical problems, but also the key to solve the difficulties of language, and according to what criteria to operate. In order to achieve national text on a higher starting point for encoding information processing platform, the paper took a Unicode international coding standards and norms set with in the codeareaYiA000 ~ A4FF expanded to 3785 characters ancient Yi characters (including 2585 words Yunnan norms and 1200 words Sichuan YiYi specification), Fig. 1 shows the ancient southern Yunnan Yi digitally encoded information processing platform needed to resolve the key theoretical and production process.

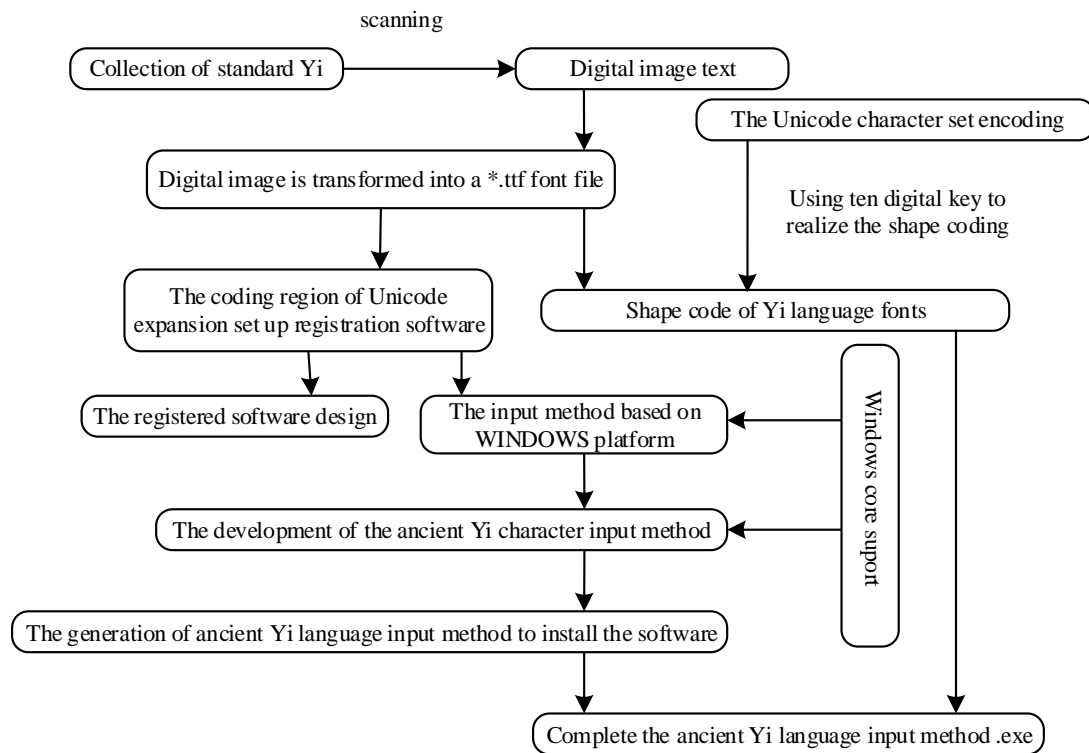


Figure 1. The ancient southern Yunnan Yi digitally encoded information processing platform needed to resolve the key theoretical and production process.

(1) Solve international coding standard Unicode area YiYi exclusive area and several other information processing platform (e.g., Microsoft Windows Vista Yi language information processing platform, Founder Unicode Yi, Yi Southwest University for Nationalities Unicode information processing platform: A before Yi Information Processing platforms, etc.) can only be placed 1220 Sichuan Yi specification text problem, this will be expanded to 3785 Yi font text (including word 2585 and 1200 word norms YiYi Yunnan, Sichuan specification), to meet the needs of ancient literature published.

(2) Expansion set for Unicode encoding region, designed a Windows, XP operating system can recognize the file by default, which is named for the registration software, Windows, XP operating system, once installed the software on this registration, the system will automatically default

Unicode encoding area expansion is set to address the phenomenon will occur after the system installation information processing platform garbled.

(3) Ancient Yi TrueType font based front-end display Unicode (ISO / IEC10646 / Unicode BMP, computer unified approach to a variety of text encoding, Unicode represents the future development trend of text encoding, this paper implements the Unicode encoding standard) international standard code design.

(4) Using a simple and efficient way to solve the input (rules are simple, not rote, ten minutes can learn) quick stroke input method, which uses barcode split, digital (numeric keys) combination of input and freedom split barcode coding input method. Solve the different levels, but not the pronunciation, users do not understand the use of the ancient Yi characters.

(5) Troubleshooting Windows operating system digitally encoded information processing platform common problems.

## Planning and design platform

### Yi ancient text and font selection production

Firstly, collected extensively distributed in Yunnan, Sichuan Yi word can be used to publish specifications (including the 3785 South Yunnan Yi (containing more than two thousand five hundred and eighty-word specification Yunnan Yi and more than twelve hundred words of Sichuan specification Yi), based on the use of computer design through word manuscript, scanning input, the program fits, use photoshop CS software for image processing, artificial modification of several steps, quality inspection, organize fonts, font testing, installation, etc., font made into a printed (in bold) and handwriting, combining ancient materials and ancient Yi Yi characteristics, using computer software (photoshop CS, font creator program4 - FCP4) developed into three southern Yunnan Yi ancient text of the original graphic treatment True Type fonts and Unicode character encoding generated code tables, and in these three text (font), and present the collection of about 3785 words of the ancient Yi character set (containing 2585 words Yunnan Yi and 1200 specification Sichuan Yi word specification), after repeated several links above, made good Yi text font is installed in the Windows system, and Yi input method developed in this paper commonly used in Fig. 2 for the use of software Yi FCP4 text font glyphs were modified, size settings, position adjustment.

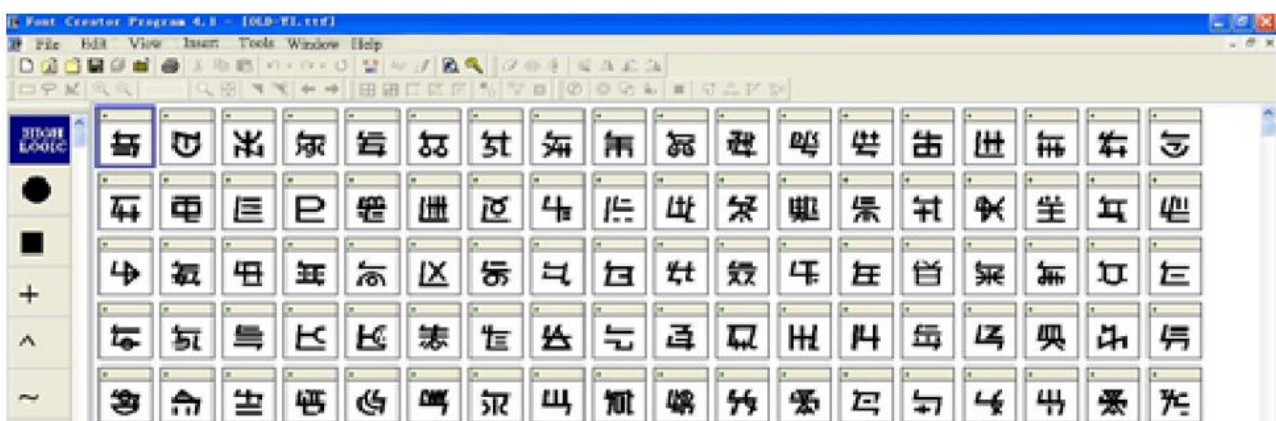


Figure 2. After the completion of installation of the Yi language fonts.

### Yi set text library code and outer code area

All collected Yi text-based digital health and freedom Split mode Yi input method is based on freedom of encoding rules are created using the program to create an external encoder input file, that external code list file, strive to be efficient, simple, easy to learn, re-code rate.

On the inner code encoding scheme. Over the years, computer commonly used to represent the ASCII characters (letters, numbers, punctuation, and control characters). With this encoding to

represent English characters is not a problem, but said that Yi, it must be expanded. Unicode-based system allows the use of 65,000 different characters on their own, enough to cover all the letters in various languages, plus thousands of symbols. According to the coded character set by relying on the existing architecture and code setting conditions within specification Yi (Yi text area within the A000 ~ A4CF, housed in southwestern China Yi text and root). Taking into account the system's scalability, standard font selection Yi Unicode character set.

Table 1. Code word mapping.

Alphabet coding region	0000.....33FF
Chinese characters extended coding region	3400.....4DB5
Chinese characters coding region	4E00.....9FA5
Standard Yi language coding region (Sichuan)	A000.....A4C6
Han Wen coding region	AC00.....D7A3
UTF-16 agent region	D800.....DFFF
User self-defined coding region (Yunnan standard Yi language)	E000.....F8FF
The restriction of the use of district	F900.....FFFF

According to the now popular characters internal code design, a character's internal code consists of two or more of the eight characters constitute a Yi calculated in accordance with the two eight characters of text. Through a lot of experiments and demonstration, this article will be placed in the text 3785 YiYi text area of international standards within theA000 ~ A4CF (housed in southwestern China Yi text and outer root word 1200), Yi remaining text placed in the "User-created characters encoding zone" in the "E900-F8FF" range as this article is assigned to southern Yunnan Yi Unicode text encoding range. Shown in Table 1, we will be within the range of coding and written on standardized coding and mapping Yi, the main tools used FCP4 UltraEdit-32 software and software, to ensure that each word corresponds to a unique specification Yi coding, This phenomenon does not occur text overlay, which can be achieved using the same input switching Yi freedoms of other input methods.

### Code word mapping

In this paper, FCP4 software specification Yi Unicode character encoding range with one mapping code. As can be seen from Fig. 3, the first specification Yi Unicode text code from E900 starts stipulated that before Yi coding interval (E900-F8FF).

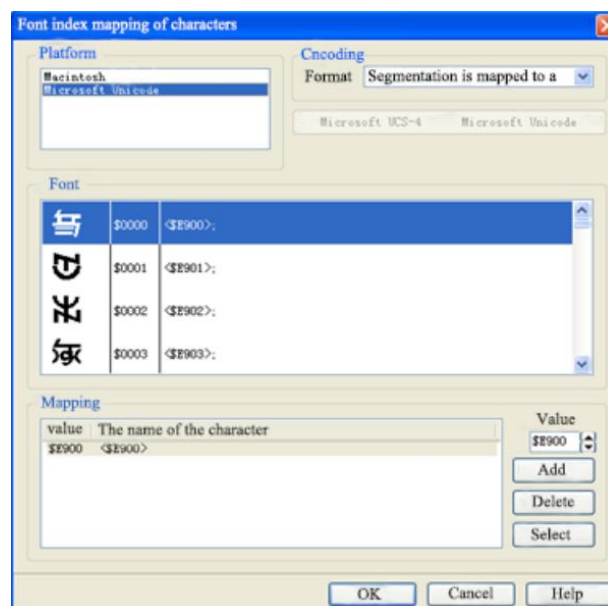


Figure 3. International Unicode encoding scheme-code word conversion.

### The Unicode encoding scheme-international code word conversion

By U-editor32 Unicode encoding software to generate Yi character code table. The font and code table for Yunnan, Sichuan Yi word specification, enter printed word. The character strokes crisp and clear font Bezier curve is very smooth, it retains the character of the ancient Yi rounded, rarely obvious triangle. Made for further input and subsequent ancient Yi typesetting, printing are very favorable, and with the input method to achieve better results. In the production process of the font by adjusting and setting the page kerning distance, to avoid overlapping phenomenon of between words. The character in the Unicode encoding (within coding-mapping) process, using a font creator program4 - FCP4 font software. Unicode coding region (within the coding region) is set to consider the characteristics of its character with a large character set, and consider scalability, the Yi text set in a user-defined area (E000-F8FF, can put 6400 words, total discharge of 2585) and part of the specification Yi district (A000 ~ A4CFh, can put 1222 words, put a total of 1200) placed the 3785 Yi text. These three characters (font) and the corresponding three input methods, once installed, can be used in more than Windows95 operating system dedicated Yi input method.

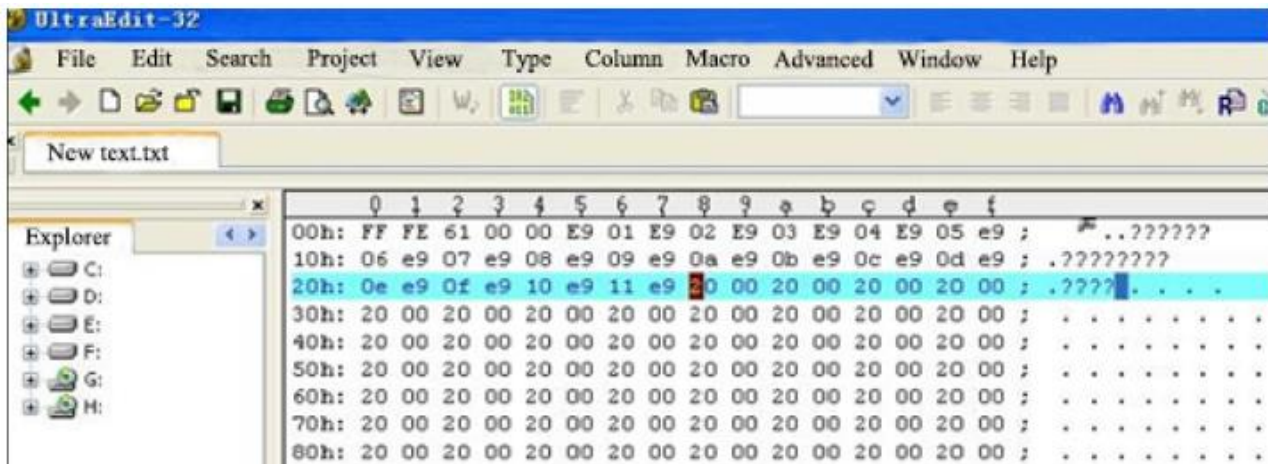


Figure 4. EditUnicode in the UltraEdit-32 HEX mode.

After Yi Unicode character encoding gone correspondence with the set up, you need to set all specifications Yi good character within the code being exported to a text document, which requires the use UltraEdit-32 software is an internal code editor, and then scheme under the Freedom of encoding rules for each character is encoded specification Yi outer code.

First, create a text file, save it in the format of Unicode format. Then, UltraEdit-32 software to open the document and interface switching to HEX mode, in this mode, the cursor at the beginning of the input character hex a decimal encoding specification Yi corresponds to the first character. It is worth noting that at the time you need to enter the code before and after the two were a coded exchange, such as "𐄂" corresponding Unicode character encoding is "E900", then in UltraEdit-32 in HEX mode, enter "00E9". Fig. 4 shows the pattern followed in the HEX editor specification Yi corresponding Unicode character encoding.

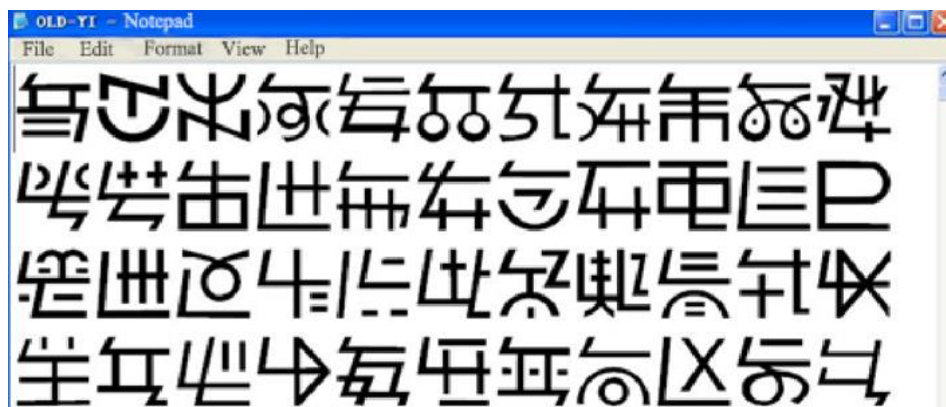


Figure 5. Standard Yi word appeared in the document of Unicode format.

When all Unicode codes are edited, the document is saved. Yi word we need can appear by modifying fonts in the document. As shown in Fig.5:

So far, the mapping relationship has been established between the inner and outer code of Yi word. And every standardized Yi has a corresponding Unicode encoding. The next step is to select a suitable coding scheme, for editing and setting of the outer code word for each of specification Yi, to achieve connect of the inner and outer codes.

### **Digital encoding scheme**

We use the barcode encoding in the paper. The user can use it to output Yi word whether people could know Yi word or use Yi word in the first time. This unique encoding is a kind of number key encoding and free split encoding which it combines the features of Yi word and writing specification.

Although the use of the frequency and extent of regulation about Yi word are far less than Chinese, Symbol system of Yi word have obvious differences with Chinese, structure of Yi word has its own characteristics.

Digital key encoding is a kind of split method. Because Yi word's stroke is simple, it often can split Yi word into three characters. Secondly, Yi word written procedure is stressed from top to bottom, and it is also streamlined from left to right. Yi word as a whole compliances with writing laws. Finally, Yi word's writing emphasizes strokes smooth and arc is mellow and beautiful. Because of the above characteristics about Yi word, the paper adopts a kind of coding scheme based stroke. For Yi word to be processed, we performed editing the outer code and distribution of the code bits of internal code.

For freedom split encoding, it is kind of barcode encoding scheme based on freedom Yi split mode, referred freedom coding scheme. Freedom coding scheme is a kind of coding mode without root that it is based on "xingtuo" rule (Some radical of specification Yi characters and letters is similar) to deal with the freedom to split specification Yi character and input. Due to the special nature of its structure specification Yi, the radical of specification Yi character is possibly similar with more letters. So, in the course of the same specification Yi characters to split, different people may have different splitting results. Therefore, in the encoding process, the different splitting results, which the same specification Yi character is split, are encoded. And in the input process the user can freely input according to their splitting results. This freely splitting mode reflects a great deal of flexibility in the input mode.

### **Draw input software of ancient Yi**

#### **Use Peg way that windows provides to achieve core technology of IME**

To use function, which is provide and supports multi-language, will make the mixed treatment of standard Yi and Chinese brings great convenience. The only thing we need to do is to analyze regional arrangements of standard Yi font, create a standard Yi font library, monitor keyboard

messages to achieve the standard input function Yi, monitor keyboard message and achieve the function of standard input. Before the input event, this mechanism could intercept the input event reaches the application and convert the action message of input devices, such as keyboard, mouse and so on, to Yi string message. Then, the Yi string message could be passed to the application window. The advantage of this mechanism is the use of system-level functions acting on all application windows and is a strong adaptability.

### Manager of input method achieves program of input method

Use Windows IME interface of common techniques to achieve Yi IPA input method by programming. IMM, for developers, allows developers to design new input method added to the system. So, Microsoft provides a set of toolkit to develop input method for the development. It is a kind of convenient and norm simple mentation that using manager of input method provided by Windows and using IME API (Application Programming Interface) which programs input method could convert the input code string to Yi string according to certain encoded rules and achieve input method.

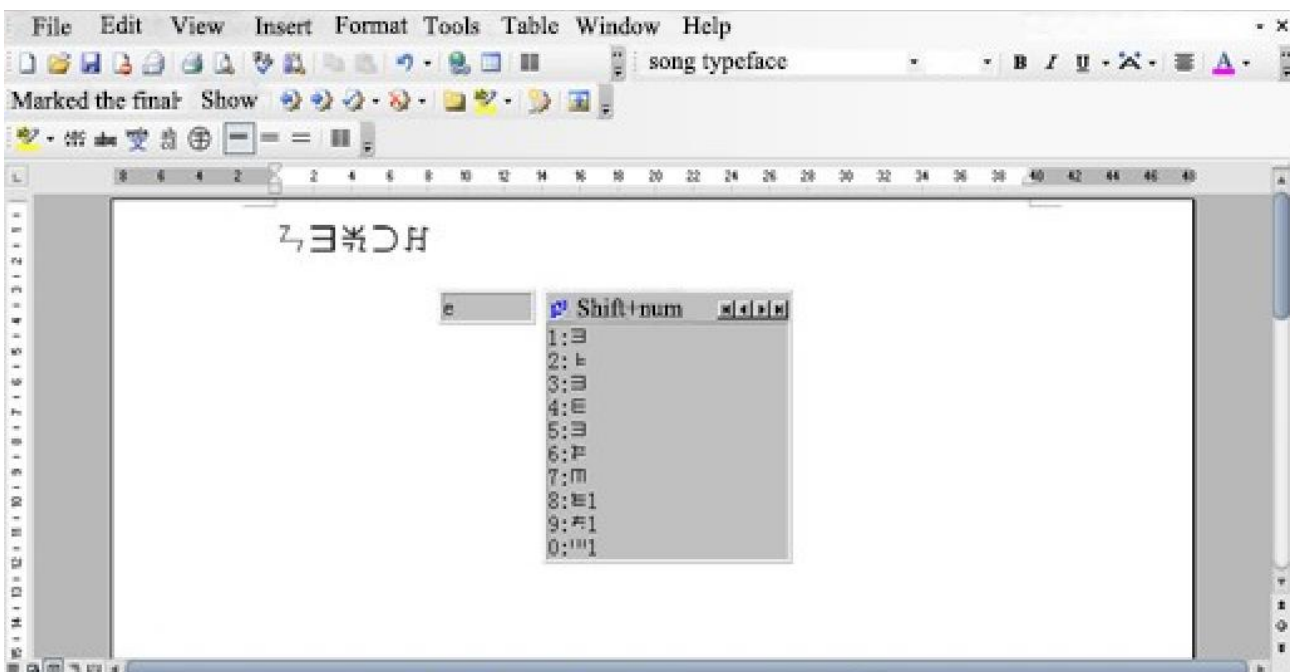


Figure 6. The realization renderings of Yi freedom split coded information processing platform in Word2003.

In the development process, design three new southern Yunnan ancient Yi digital encoded information processing platform based on the characteristics of ancient Yi characters in this paper such as the realization renderings of Yi freedom split coded information processing platform in Word2003 in Fig.6 Using this platform allows users, which do not understand Yi and pronunciation, to complete entry work of Yi string. And it provides convenience for ancient development and publication and distribution of Yi literature. It is characterized by its fast input resistance, being easy to learn and flexibility. Therefore, the program has a stronger feasibility and practicality, which has a more extensive use value for ancient Yi printing, office automation and information processing.

At present, the test version of digital information processing platform series has been developed. They have been provided to the cooperative units to try. They receive a high evaluation in cooperation with Yunnan Honghe National Institute. Three kinds of digital information processing platform are tried in Yunnan Nationalities Publishing House which gives a higher rating.

## Summary

Currently, Yi character input methods has been made significant progress in Yunnan, Sichuan, Guizhou, Guangxi and other places, especially 1165 Sichuan standard Yi language character and 688 Graphic input method of character created by Sharma La Ki, has been used in Yi language writing, computer software development, publication of the Yi nationality books as well as other materials related to the documentation. But these Yi character input methods can not cover treasure of national culture the Yi nationality ancient characters and meet the needs of ancient literature publishing. In this paper, the innovation for the development of the Yunnan ancient Yi digital information processing software platform and made up for the shortage, the better inheritance of Yunnan minority language and literature the distinctive cultural flower, to the benefit of future generations.

## Acknowledgement

This work was financially supported by the National Natural Science Foundation of China (61363085), Major projects in Yunnan Province (ZD2013013), High level construction of university scientific research project of Yunnan MinZu University. State Language Commission authorize research projects, WT125-61.

## References

- [1]Wang JiaMei.Yi Ancient Digital Research - Yi Ancient special website.College. 2006.07. 149-150.
- [2]Li XunShan, Wang JiaMei (Corresponding author).Research on Yunnan Standard Yi Language Fonts Design and Character Coding ,Electronic Science and Technology, 2011.24(5) . p97.
- [3] FENG Hao,WANGHui,WANGJia-mei (Corresponding author). Design and implementation of Yi character input method based on free split mode. Journal of Computer Applications. 2010. A01, 306-308.
- [4] FENG Hao, ZHOUYing,WANG Jia-mei. Popularization and Application of the Yi Character Input Method's Free Split Mode. Modern Computer,2011.9.3-5, p11.
- [5] CHEN Shun-qiang, ZHANG Yang, XIONG Jian.Sichuan ancient Yi language fonts design and character repertoire code. Journal of Southwest University for Nationalities (Natural Science Edition), 2009,35(4):913-918.
- [6]PU Zhang-kai. A Dictionary of South Yunnan Yi. The Nationalities Publishing House of Yunnan. 2005.10.
- [7]Yi character set of Yunnan,Sichuan, Guizhou and Guangxi. Published jointly by three Nationalities Publishing House of Yunnan, Guizhou and Sichuan. 2001.
- [8]ShaMaLayi. Development of computer operating system of the Yi language [J]. Journal of Southwest University for Nationalities (Natural Science Edition), 2003.29(1):1-8.
- [9]ZHU Jian-jun. Some Thoughts on the Setting up an Ancient Yi Character bank. Journal of Huzhou Teachers College[J]. 2003.25(1):22-25.
- [10] Li Jinfa. Exploitation of Yi Character Coding Conversion Software. Journal of Yunnan Nationalities University (Natural Sciences Edition), 2008.14(1)
- [11] Fu Xianghua. Design and implement theIOB operation platform and IOB database system. Master's thesis. Northwest A&F University. 2002.



[12] Fang Ziwen. Design and implement the Ancient Yi Windows input platform. Master's thesis. Yunnan Nationalities University. 2008.5.

[13] Lu Jixing, Jiang Weike, Zang Yueli, Qu Bin. Input method based on Windows IME[J] Journal of Agricultural University of Hebei, 2003.26(5):290-292.