

# NaXi Pictographs Input Method and WEFT

Hai Guo

Dalian Nationalities University, Dalian, China  
 Email: guohai@dlnu.edu.cn

Jing-ying Zhao\*

Dalian Nationalities University, Dalian, China  
 Email: dalianzjy@gmail.com

**Abstract**—NaXi Pictograph , which is the only hieroglyph in use is important for researching into the evolution of the character. In the past, processing NaXi Pictograph by hand, that is very inefficient. The NaXi Pictograph information processing modules is developed, such as pictograph Outline Font lib, the input method module, the embedded module and so on. A method of NaXi Pictographs Outline font extraction is proposed, which has been testified better in the Linux. As the specialty of NaXi Pictographs,two input methods -- NaXi pictographs pinyin and graphic primitive input method are advanced, by evaluating ,which is found that the second is better than the first in precision. For display in web, The web embedding fonts technology of NaXi Pictographs is brought forward, and The web embedding fonts technology of NaXi pictographs makes Internet clients browse NaXi pictographs web without downloading NaXi pictographs font. The development of the NaXi Pictograph Information Processing System is significant to research into NaXi pictographs and apply it.

**Index Terms**—NaXi pictographs, information processing, outlines font, IME (Input Method Editors), embedded font, WEFT

## I. INTRODUCTION

NaXi pictograph belongs to the NaXi language of yi language branch of Tibetan-Burmese languages which was created by the ancestor of NaXi. It’s credited as “the only living ancient hieroglyph” and still being used in writing lection and composition and in the field of communication, therefore, NaXi pictograph has a special role in the history of human character. Since hieroglyph appeared in the early stage of the development of characters, through a deep study of it we can get something about the evolutionary history of human characters and human culture which will make a great contribution to research on the origin of modern characters. Many experts and scholars at home and aboard have been working on NaXi pictograph for a long time, among which Harvard and Yunnan Academy of Social Sciences lead a dominant role. But the problem that a lot of literature and ancient books can’t be processed efficiently makes it urgent to realizing an information processing system for NaXi pictograph.

The traditional hand-drawing method of processing NaXi pictograph has low efficiency and can’t guarantee the problem can be solved by the output of computer’s

standard font, therefore, our project designs and develops the NaXi pictograph information processing system for the first time which ends the processing history of NaXi pictograph without computerization.

## II. NAXI PICTOGRAPH PROCESSING SUMMARIZATION

The Naxi Pictograph input platform is developed for special groups, including units of printing and publication, Naxi pictograph research institutes and art designing companies, so it’s not a universal software platform. These reasons make it definite that our development must meet the requirement of different customers. The analysis of application level of this project can be summarized as Table I.

According to the analysis above, our project has designed various Naxi pictograph input processing systems which can meet specific requirement of different customers. For professional customer doing publication and printing, we developed Naxi pictograph standard edition with TrueType and PostScript outline fonts, and Four input methods including internal code input method, Latin input method, graphic primitive Method and English input method, which can fully satisfy the needs of printing and typesetting. As for the Naxi pictograph research institutes, we developed Naxi pictograph TrueType outline font and Naxi Pinyin input method. Considering that the artistic designing units just need the font style outline, their system only have Naxi pictograph TrueType outline fonts and Naxi pictograph English input method. At last, we developed web embedding fonts and PDF document creating technology of Naxi pictograph

TABLE I.  
 THE ANALYSIS OF CUSTOMER REQUIREMENT OF NAXI PICTOGRAPH INPUT PLATFORM

Regular	customer requirement
customer group	high accuracy, fast input, output, various convenient input method
Printing Publication	general output accuracy, advanced input method
research institutes	various of output font styles, input method easily to learned
artistic designing	offer web pages and electronic document browsing
ordinary user	customer requirement

\*Corresponding author: Jing-ying Zhao Email: dalianzjy@gmail.com

for ordinary customers.

### III. THE PRINCIPLE OF NaXi PICTOGRAPH INFORMATION PROCESSING

#### A. The principle of NaXi Pictograph Outline Fonts

Since outline font can be zoomed in and out at will without distortion, it is widely applied in printing, typesetting, character processing and art creating, and Windows we uses everyday is displayed with outline font too. This type of font is composed of Bézier curve, and the TrueType font uses quadratic Bézier curve, the PostScript font uses cubic Bézier curve, so the reducibility of PostScript font is better than the TrueType font. A Bézier curve is controlled by three points, as is shown in Fig. 1, when the middle control point changes the whole curve changes too. If an outline word consists of  $n+1$  points ( $P_0, P_1 \dots P_n$ ), it needs  $m$  Bézier curves to form the font. Equation (1) shows the detail.

$$m = \sum_{i=0}^n \binom{i}{n} t^i (1-t)^{n-i} P_i \quad (1)$$

A PostScript outline is formed of a group of cubic Bézier curves which can be described in (2), (3):

$$x = a_x * t^3 + b_x * t^2 + c_x * t + d_x \quad (2)$$

$$y = a_y * t^3 + b_y * t^2 + c_y * t + d_y \quad (3)$$

A TrueType outline is formed of a group of quadratic Bézier curves and every curve is defined by three control points. For a quadratic Bézier curve's three control points are  $(A_x, A_y)$ ,  $(B_x, B_y)$  and  $(C_x, C_y)$  described in (4) and (5):

$$P_x = (1-t)^2 A_x + 2t(1-t)B_x + t^2 C_x \quad (4)$$

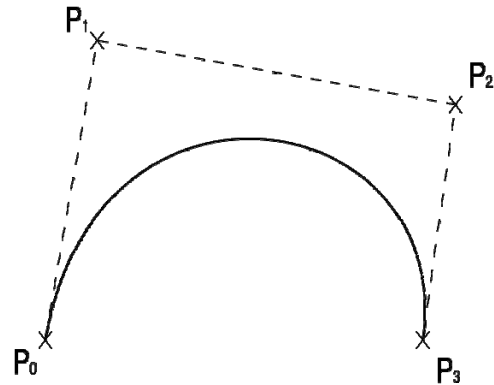


Figure 1. Bézier Curve

$$P_y = (1-t)^2 A_y + 2t(1-t)B_y + t^2 C_y \quad (5)$$

By modifying the parameter  $t$  from 0 to 1, all values of  $p$  defined by  $A$ ,  $B$  and  $C$  can be generated, from which the quadratic Bézier curve is obtained.

Font is defined as a combination of a set of outline curves and every curve includes three or more points. The illustration using Bézier outline curve of NaXi pictograph "deer" is shown in Fig. 2. The outline is constituted of a set of instructions which pictured the character's exterior.

All the parameter information of outline fonts is stored in a NaXi pictograph table "glyf". Because of the high complexity and similarity of NaXi pictograph, the precision of outline curve delineation is greatly higher than the dot matrix font delineation. Based on the two outline font technique, we developed NaXi pictograph TrueType font and PostScript font which basically met the requirement of processing basic model of NaXi pictograph with computer.

#### B. The Feature Extraction Method of NaXi Pictograph Outline Font

The outline font is filled with a series of outline curves and delineating the outline accurately is the key point for the success of a font. It can be seen from



Figure 2. the Outline Curve delineation of NaXi Pictograph "deer"

Figure3 that the NaXi pictograph is complicated, various and also with many strokes which make it difficult for truly reducing the outline of NaXi pictograph. According to these characteristics, our project proposes a dual-mode transformation algorithm for extracting the NaXi pictograph outline points and uses some rules to check whether a pixel point is in the outline.

Rule1: if the center of a pixel point is in the outline, the point is lightened and becomes a part of the outline curve.

Rule2: if an outline line passes the center of a pixel, the pixel is lightened.

Rule3: if a scanning line located in the center of two adjacent pixels (horizontal or vertical) intersects with the on-Transition and the off-Transition and the points on this line haven't lightened by rule1 and rule2, the left endpoint is lightened when this line is horizontal, while the right endpoint is lightened when it's vertical.

Rule4: rule3 can be used only in the case that two contour surfaces still intersect the scanning line in two-way, but this doesn't mean these pixels are "stubs". By checking the scanning line-segment formed a square with crossing scan line-segment, whether they intersect each other through two contour surfaces can be verified. There is the possibility which is small but exists that more than one contour intersect with discontinuity point, so it's necessary to control some characters outline using grid-fitting.

The method discussed above is used in the secondary development of Pgaedit based on Linux which effectively prevents the generation of discontinuity point, and the accuracy of outline point extraction of NaXi pictograph reaches 99.99%.

#### IV NaXi PICTOGRAPH FONT DEVELOPMENT

A computer font is an electronic data file containing a set of glyphs, characters, or symbols such as dingbats. Although the term font first referred to a set of metal type sorts in one style and size, since the 1990s most fonts are digital, used on computers. There are three basic kinds of computer font file data formats:

- Bitmap fonts consist of a series of dots or pixels representing the image of each glyph in each face and size.
- Outline fonts (also called vector fonts) use Bézier curves, drawing instructions and mathematical formulas to describe each glyph, which make the character outlines scalable to any size.
- Stroke fonts use a series of specified lines and additional information to define the profile, or size and shape of the line in a specific face, which together describe the appearance of the glyph.

Bitmap fonts are faster and easier to use in computer code, but inflexible, requiring a separate font for each size. Outline and stroke fonts can be resized using a single font and substituting different measurements for components of each glyph, but are somewhat more complicated to use than bitmap fonts as they require additional computer code to render the outline to a bitmap for display on screen or in print.

The difference between bitmap fonts and outline fonts is similar to the difference between bitmap and vector image file formats. Bitmap fonts are like image formats such as Windows Bitmap (.bmp), Portable Network Graphics (.png) and Tagged Image Format (.tif or .tiff), which store the image data as a grid of pixels, in some cases with compression. Outline or stroke image formats such as Windows Metafile format (.wmf) and Scalable Vector Graphics format (.svg), store instructions in the form of lines and curves of how to draw the image rather than storing the image itself.

A bitmap image can be displayed in a different size only with some distortion, but renders quickly; outline or stroke image formats are resizable but take more time to render as pixels must be drawn from scratch each time they are displayed.

Fonts are designed and created using font editors. Fonts specifically designed for the computer screen and not printing are known as screenfonts.

#### B. NaXi pictograph font library

Font library is the basis and also the core of the system, accurate font and correct code set solid foundation for developing the complete system. The NaXi pictograph font library includes NaXi pictograph TrueType outline font and NaXi pictograph PostScript font. TrueType outline font is composed of quadratic Bézier curves and PostScript is built with cubic Bézier curves. Since a quadratic Bézier curve can be transformed to a cubic Bézier curve, we develop TrueType outline font first and then realize the transformation by expert tools.

The development of NaXi Pictographs Outline Font is mostly going by to design character draft, to scan the font, to digitize the font, to modify and coordinate the font, to examine the quality of character, to conformity character storeroom. After calculation, draft is scanned as lattice character storeroom in high precision by scanner, at the same time the character storeroom code is given. To simulate and conformity digitalization is according to the method of double pattern switch, which tries its best to shift lattice graph to numeral information as real as manuscript automatically (curve outline). The outline point, line, angle and location can be rectified through parameter controlling, which present extraordinarily important in the process of producing NaXi Pictographs with complex font and large diverse manner.

The outline font development can be divided into four steps: font and script designing, digital fitting, modifying and font generating. Our project uses the original script of A dictionary of NaXi pictograph sound-indication which is transformed to standard bit map by scanning, and then the describing points are extracted for digital fitting, after that the NaXi pictograph outline font is integrated and created. This kind of technique can guarantee the veracity of original script and the generated outline font has the features of excellent reducibility and vector property as well.

#### V NaXi PICTOGRAPH INPUT METHOD BASED ON IMM-IME

TABLE II. NAXI PINYIN CODE

IPA	Latin letter	IPA	Latin letter	IPA	Latin letter	IPA	Latin letter
p'	p	p	b	b	bb	f(w)	f(w)
t	d	t'	t	d	dd	n	n
l	l	k	g	k'	k	g	gg
ŋ	ng	h	h	tʃ	j	tʃ'	q
dz̥	jj	n̥	ni	ç(z)	x(y)	z̥	r
ʃ	sh	dz̥	rh	tʃ	zh	tʃ'	ch
z	ss	s	s	dz	zz	ts'	c
ts	z	i	i	u	u	y	iu
a	a	o	o	ə	e	v	v
u	ee	ər	er	e	ei	æ	ai
ie	iei	iæ	iai	ia	ia	iə	ie
uei	ui	uæ	uai	ua	ua	uə	ue

A. IMM-IME Introduction

As Windows is the most widely used operation system at present, this paper focuses on the input system for NaXi pictograph based on Windows. The input method based on Windows transforms standard ASCII string into

NaXi word or string using some particular coding rule. With different application program the user can not design a transformation program himself, due to which the task of inputting NaXi pictograph should be taken by the Windows system administration.

As shown in Fig.4, at first, the keyboard event of NaXi pictograph input system is received by the Windows file use.exe, then use.exe transfers the event to the Input Method Manager (IMM), after that the IMM conveys the event to the input method editor which translates the keyboard event to its corresponding NaXi character (or string) with reference to user's encoding dictionary, when this is done the translated event is propagated back to use.exe and then to the executing application, until now the whole input process of NaXi pictograph is finished[4-5].

B. The Basic Structure of NaXi Pictograph IMM-IME

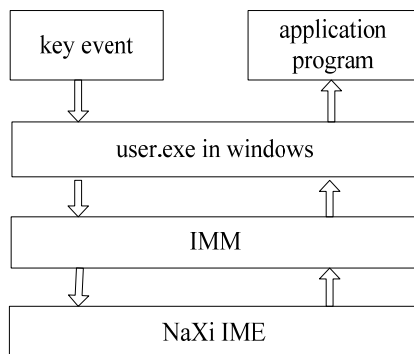


Figure 3. the Input Method Principle of NaXi Pictograph.

The IMM-IME structure provides various input methods for applications, each thread of an application can keep an active input window. The processing order of other messages won't be disturbed by inserting the NaXi pictograph message to message circle. The head file

immdev.h should be included for using these new features. The detailed working principle of NaXi pictograph Pinyin input method is shown in Fig. 3.

C. NaXi pictograph pinyin input method

Current input method includes two types: Pinyin and font. The font input method can be divided by strokes which needs the user has the ability to write. Unfortunately, writing NaXi pictograph is so hard for the ordinary user that makes the shape code unsuitable for NaXi pictograph being inputted to the computer. The phonetic code input method just requires the user know how to read the character, therefore, Pinyin input method is more suitable for NaXi pictograph.

Early dictionaries of NaXi pictograph sound-indication employ International Phonetic Alphabet to mark NaXi character, while the computer uses Latin letters as input code, the conversion between NaXi Phonetic and Latin Pinyin becomes necessary. Table II lists the mapping between International Phonetic Alphabet and Latin Pinyin in detail. When designing the input method of NaXi pictograph one can encounter the problem of too long encoding due to the characteristics of NaXi pronunciation. For instance, the NaXi character 𐄎𐄎 can be coded as "ssoxiqssoddassa", it needs fifteen English characters to map this single one character which will result in very low efficiency when input an article. Research shows that the initials repetition phenomenon is common in NaXi pictograph pronounced coding, through the method of designing code with simplified initials the Pinyin input method of NaXi pictograph can greatly reduce the coding length and improve the coding efficiency. Let's take the character 𐄎𐄎 as an example, its pronounced coding is "ssoxiqssoddassa", after the simplification of pronunciation it becomes soxiqsodasa, the length reduces to eleven from fifteen. NaXi pictograph consists of 2120 characters and the average coding length is twelve bits which is reduced to eight bits after the simplification of pronunciation, so the input speed is also increased.

D. NaXi pictograph graphic primitive input method

1)the Structure Coding Method of NaXi Pictograph: The NaXi Pictograph has four common structures, those are undivided whole structure coded by b, up and lower structure coded by s, right and left structure coded by z, surrounding structure coded by b. The NaXi Pictograph character of undivided whole structure are such as 天, 鸟, 虫, 山, 水 and so on, the upper and lower structure, 采, 采, 采, 采, 采; right and left structure, 采, 采, 采, 采, 采; surrounding structure, 采, 采, 采, 采, 采. After the coarse segmentation and coding above, the fine granularity coding is introduced.

2)the Graphic Primitive Coding Method of NaXi Pictograph: The NaXi pictograph character hasn't radical component such as Chinese characters, so the representation method using Graphic Primitive is proposed in this paper, which inspired by syntactic pattern recognition. The basic elements of graph, graphic

curve, oval curve and rectangle. As showed in Table III, the basic graphic primitives are coded. Different with Chinese and other characters, many NaXi pictograph characters contains digit, for example 采 (dice), the number of dots, and 采 (fire), the number of vertical curve, and 采 (treasure), the number of circle. So the quantity is coded, one is 'y', two is 'e', three is 's', four is 'f', five is 'w', six is 'l', seven is 'q', eight is 'b', nine is 'j', the number greater than nine is 'd'. The user interface of NaXi primitive input method is shown in Fig. 4.

E. Evaluation of NaXi Pictograph Input Method

In order to evaluate the efficiency of NaXi proposed by this paper, we develop an evaluating system for NaXi pictograph input method. Its flow procedure is shown in Fig. 6. The evaluating system employs the Windows message mechanism to automatically converse the NaXi Pinyin text to NaXi pictograph text under the conditions of activating the NaXi input method with and without optimization. After getting the evaluation result, we can calculate the conversion accuracy.

In the experiment, we choose five NaXi texts to the evaluating system, the results demonstrate that the NaXi graphic primitive input method achieved higher average conversion accuracy than the Pinyin input method. Table IV shows the details.

VI WEB EMBEDDING APPLICATION

Aiming at the ordinary customers' requirement, our project also made some study on the application of NaXi pictograph and developed Web embedding and PDF embedding technology.

A. The Classification of Web Embedding

The lattice font has been gradually eliminated at the

TABLE III. GRAPHIC PRIMITIVE CODE OF NAXI PICTOGRAPH

Graphic Primitive	code
point	a
line	b
circle	c
circular curve	f
left oblique line	g
right oblique line	h
vertical line	i
vertical curve	g
oval curve	k
rectangle	l
right oblique line	h

primitives, are point, line, circle, circular curve, left oblique line. right oblique line, vertical line, vertical

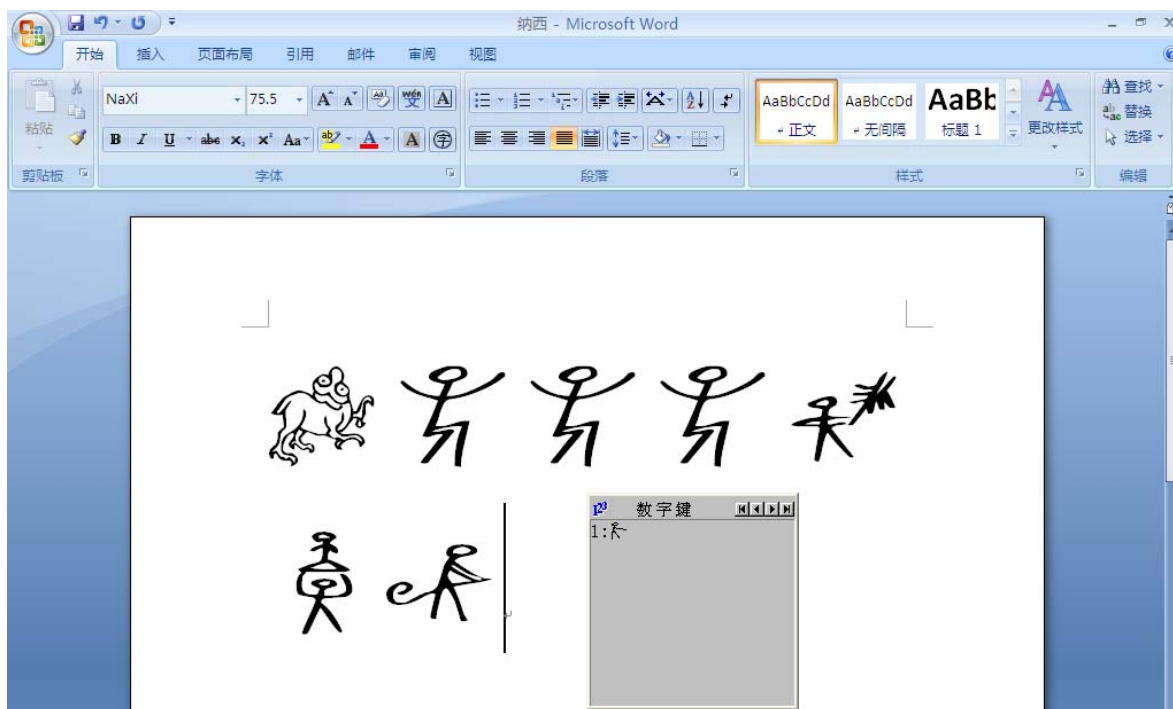


Figure 4. NaXi Pictograph Primitive Input Method.

TABLE IV.  
COMPARISON OF THE NaXi PINYIN INPUT METHODS

Conversion Accuracy	text1	text2	text3	text4	text5	Average Accuracy
Pinyin Method	99.1%	98.1%	98.4%	96.1%	97.8%	98.1%
graphic primitive Method	99.6%	99.1%	98.9%	97.1%	98.1%	98.6%

beginning of 90s in last century with the updating of operating system, the new outlines font has replaced it. The mainstream outlines font includes:

① PostScript Type 1 font, brought forward by Adobe corporation years ago, applied in publication typesetting system, belongs to the first generation of outlines font.

② TrueType font, brought forward by Apple and Microsoft. Because it has so many merits, it has been used in a variety of operating system on Mac and Pc, belongs to the second generation of outlines font.

③ OpenType font, Put forward by Adobe and Microsoft as a new generation of outlines font standard. It fuses the merits of Type 1 and TrueType, belongs to the third generation of outlines font.

The TrueType font embedding technology of NaXi pictographs can be divided into embedded OpenType and True Doc. Put forward by Microsoft, Embedded OpenType technology can compress TrueType font into Eot file, and then embeds it into HTML web pages. While TrueDoc brought forward by Netscape and Bitstream can compress TrueType font into TrueDoc file and then embed it into HTML web pages.

**B. The Principal of Web Embedding Technology of NaXi Pictographs**

Web embedding technology of NaXi pictographs uses

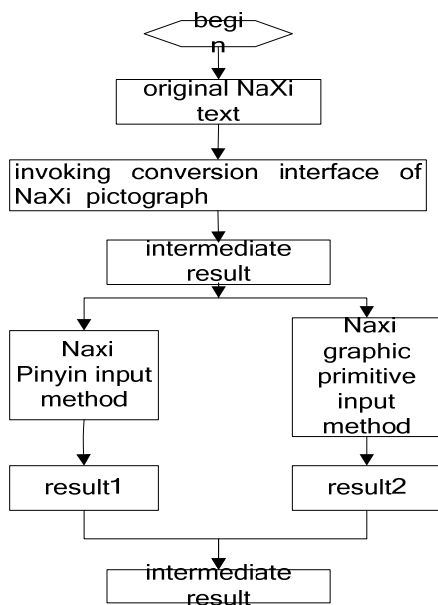


Figure 5. NaXi Pictograph Input Method.

CSS2 calling compressed OpenType font files to embed NaXi pictographs into web page. Downloading compressed fonts of NaXi pictographs downloaded to clients' temporary directories lets clients browse web pages with NaXi pictographs

correctly without installing NaXi pictographs font.. The principle is that a client sends requirement for browsing, and then HTTP server sends HTML files to client browser. Client sends the information of web pages contained NaXi pictographs to server through CSS2, the server send this information to EOT database. EOT calls the TrueType font files stored in HTTP server, then HTTP server sends this inmessage back to the browser. The NaXi pictographs will be deleted autonomously after the client closes the browser. In this way fonts copyright can be better protected from piracy. By calling CSS2 several times in a web page, not only NaXi pictographs but Chinese, Mongolian, Tibetan and other minority languages can be displayed in the same web page.

**C. Realization of Web Embedding Fonts Technology of NaXi Pictographs**

1) *The environment of developing web embedding fonts of NaXi pictographs:* The development environment is an environment of making web pages, EOT files, and CSS tables, it is mainly composed of DreamWeaver, WETF (Web Embedding Fonts Tool), Font Creator and etc. Free BSD installing Apache is adopted as the server. Firstly this structure is totally compatible with Microsoft IIS system. Secondly, it provides more useful functions, faster operating speed and better stability than Microsoft IIS system.

2) *The application of CSS2 in web embedding fonts of NaXi pictographs:* pictographs TrueType font into web pages possible. From the way that CSS is inserted, there are three kinds of CSS: inline mode list, embedded mode list and exterior mode list, and the inline mode list and embedded mode list are broadly used in web embedding fonts technology of NaXi pictographs. The key sentence of CSS is @font-face, which has defined the name, type, thickness and other information of a planted font.

3) *The generation of planted font database of NaXi pictographs:* The substance in creating font database is to compress NaXi pictographs TrueType fonts into OpenType fonts. There are many ways in generating NaXi pictographs planted databases, this paper adopts Microsoft WETF ( Web Embedding Fonts Tool). Before planting the validity of NaXi pictographs TrueType fonts in the development environment are checked by WEFT. WEFT displays the font validity in graph. .

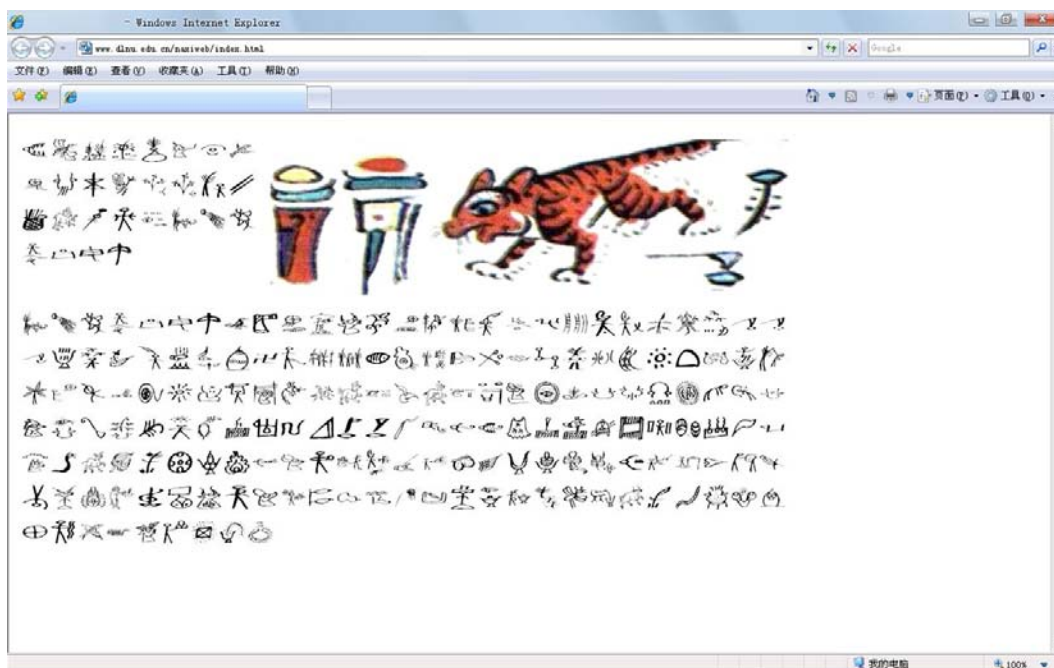


Figure 6. NaXi Pictograph WEFT Web.

After the success of font-checking, add the needed NaXi pictographs into EOT files, then calling NaXi pictographs font while browsing will not be a problem.

4) *The embedding of NaXi pictographs font:* You can input NaXi pictographs by using NaXi pictographs input method developed by the information system lab, then create a web page containing NaXi pictographs with Dreamweaver. Then insert codes as follows between <head> and </head>:

```
<STYLE TYPE="text/css">
<!-- /* $WEFT -- Created by: GuoHai
(guohai@ourcampus.net) on 2003-4-4 -- */
@font-face {
font-family: NaXi;
font-style: normal;
font-weight: normal;
src: url(NAXI.eot); }
-->
</STYLE>
```

Embedding function @font-face provides with four parameters: you can use font-family to name this font in the current webpage, this project is defined as NaXi; font-style can be anyone of normal, italic or oblique, commonly defined as normal; font-weight can be normal, bold, bolder, lighter or other legal thickness value; FontURL is a URL pointing to OpenType files, normally adopts absolute pathway. After saving the NaXi pictographs webpage, you can test it by transferring it to the HTTP server. The testing Figure can be seen in Fig. 6. Up to now, the preliminary process of planting NaXi pictographs is officially finished.

VII CONCLUSION

A complete NaXi pictograph information processing platform is developed in this paper. It ends the history of

NaXi pictograph processing without computer, provides valuable reference for creating other minority languages information processing system and also plays a significant role in promoting the computerization process of minority characters in China.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (No. 60803096).

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529-551, April. 1955.
- [2] Joseph Francis Charles Rock, *A Nakhi-English encyclopedic dictionary*, I.M.E.O, Rome, 1963.
- [3] Liu Yongkui, Guo Hai, Lu Guiyan, Li Hongyan, "Input technology and information processing of NaXi pictograph", *Journal of Computational Information Systems*, vol.3, no.1, p361-368, February, 2007.
- [4] Guo Hai, Che Wengang, Nie Juan, Li Bin etc, "Web embedding fonts technology of NaXi pictographs", *Jisuanji Gongcheng/Computer Engineering*, vol.31, no.17, pp.203-204+207, Sep. 2005.
- [5] Guo Hai, Zhao Jing-ying, "Development of the NaXi Pictographs Information Processing System", *Control & Automation*, vol.22, no. 22, pp.122-124, 2006.
- [6] Xie Qian, Jiang Li, Wu Jian, etc, "Research on Chinese Linux input method engine standard", *Jisuanji Yanjiu yu Fazhan/Computer Research and Development*, vol.43, no.11, pp.1965-1971, November, 2006.
- [7] Tseng Chun-Han, Chen Chia-Ping, "Chinese input method based on reduced Mandarin phonetic alphabet", in Proc. *INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing*, 2006, pp.733-736.

- [8] Tanimoto Yoshio, Nanba Kuniharu, Rokumyo Yasuhiko, etc, "Evaluation system of suitable computer input device for patients", Proceedings of the Third Workshop - 2005 IEEE Intelligent Data Acquisition and Advanced Computing Systems, 2007, pp. 369-373.

**Hai Guo** received the M.Sc. degree in pattern recognition and intelligent system from the Kun Ming University of science and technology, China, in 2004.

He has been teaching at Dalian Nationalities University since graduation. For many years, he has done research in

image processing, pattern recognition. He has published more than 10 papers.

**Jing-ying Zhao** received the M.Sc. degree in computer application from the Chang Chun University of science and technology, China, in 2003.

She has been teaching at Dalian Nationalities University since 2003. For many years, she has done research in image processing, pattern recognition. He has published more than 10 papers