



Displaying Multi-script Data on the Web



**Kiyonori NAGASAKI
Makoto MINEGISHI
Izumi HOSHI**

文字焼き

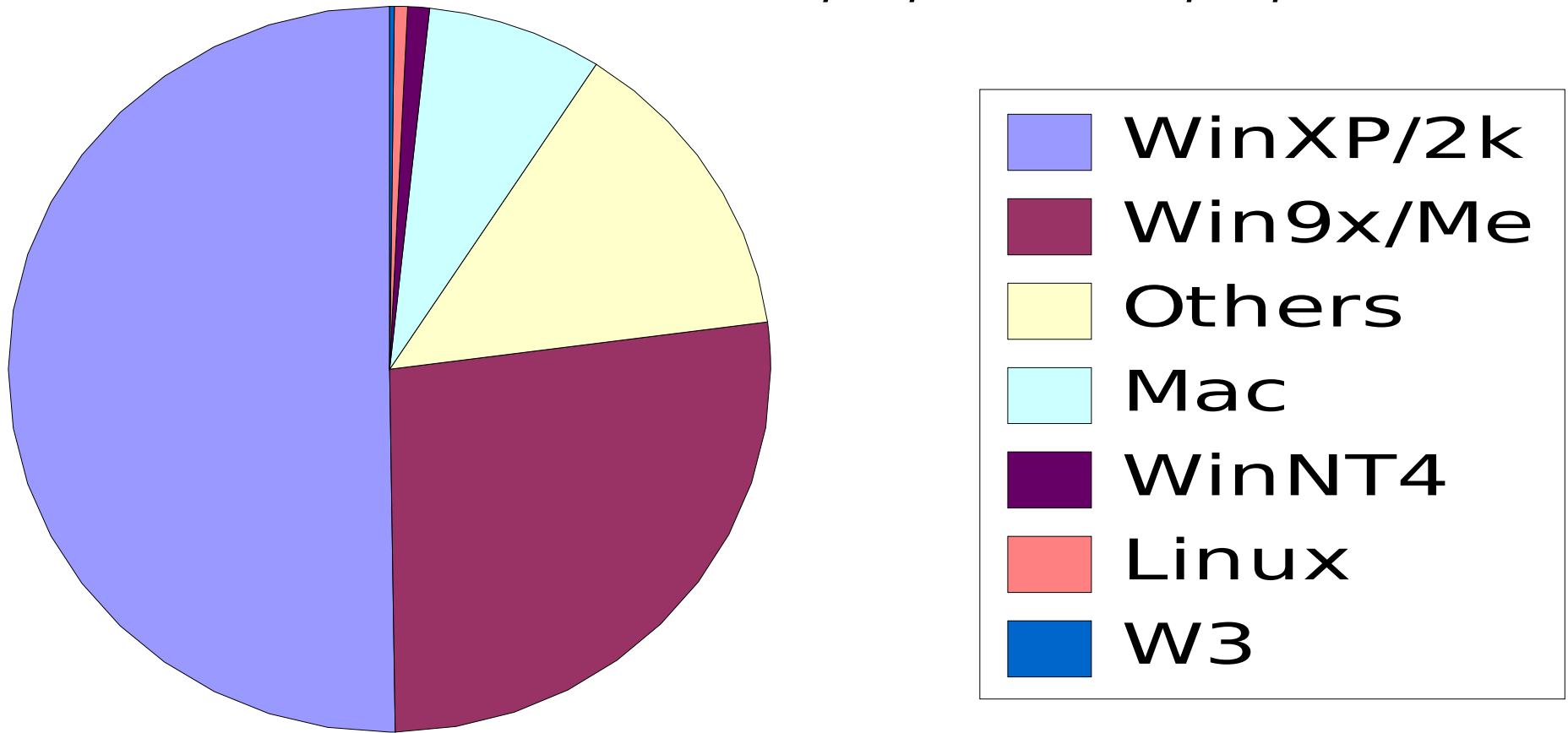
Multilingual Imager for Web

- Currently, **easy** computer environments for multi-scripts require:
 - Unicode
 - Opentype
 - ex. Windows2000/XP, Mac OS X
- However, even today these standards cannot be installed on half of all Web clients.



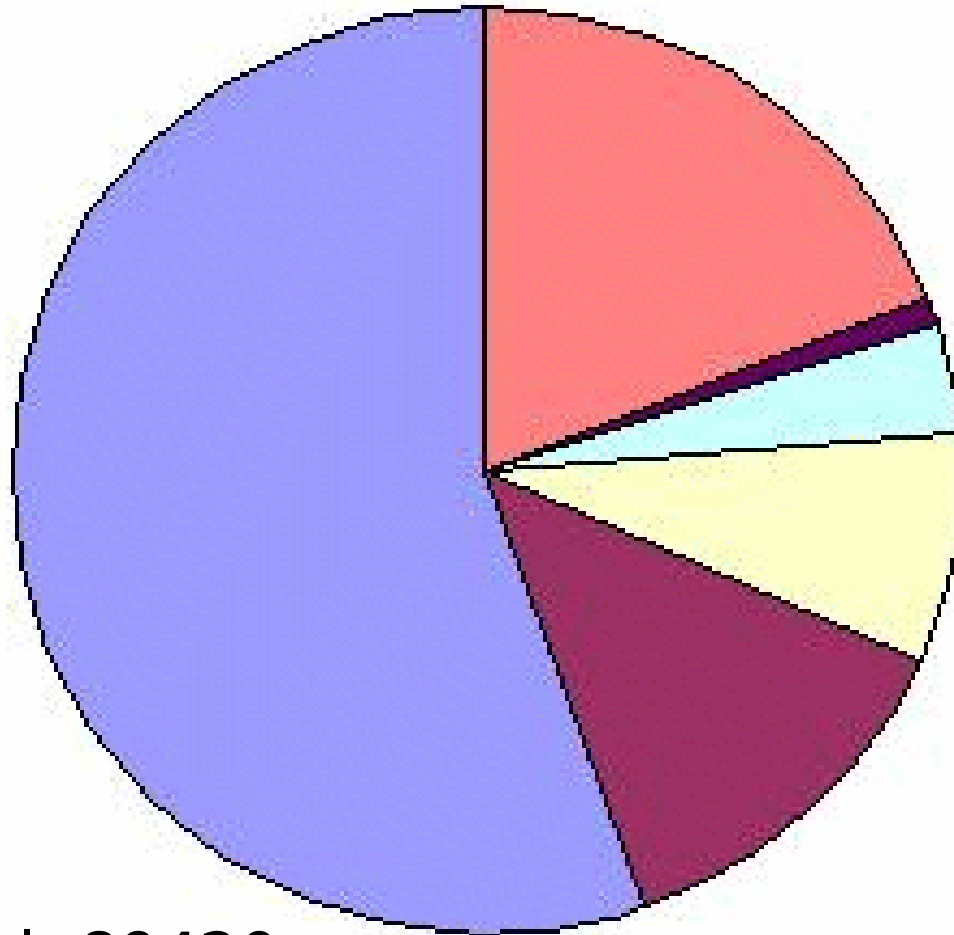
Percentage of OSs that access “Mojiyaki”

2003/10/27 - 2003/11/27



Total : 141669

Two years ago: 2001/11 - 2001/12



Win9x	55%
Win2k	13%
Mac	8%
Unix	1%

Total: 89420



文字焼き

Multilingual Imager for Web

- In order to display multi-script texts on the Web, some **improvising** techniques are needed for old clients, such as Windows9x and Me.





Examples of improvising techniques to display multi-scripts on the Web

- Install separate s for each script.
- Drawing the scripts with Java applets on the browser.
- Creating fonts as images.



文字焼き
Multilingual Imager for Web

Installing fonts for each script.

- The installation of fonts is not easy for many Web users.
- As various character codes are used on the Web, all these codes would require separate fonts.
- Problems of font copyrights
- Some Web clients do not allow font installation (ex. Web-implemented mobile phones, or PDAs)



文字焼き

Multilingual Imager for Web

Drawing scripts with Java
applets on the browser.

Java applets are too heavy for
the old clients used in many
schools or Asian countries.





Creating fonts as images.

Easy for old clients that use a graphical
Web browser.

No need for procedures, such as
installing fonts.





Features of “Mojiyaki”

Dynamically drawing fonts as images

Display as inline images on Web browsers that are capable of displaying images.





Mojiyaki can currently display these scripts.

- Devanagari (Hindi, Sanskrit)
- Tibetan
- Thai
- Laos
- Arabic Scripts (Arabic, Uighur)
- Japanese, Korean, (Chinese)





Mojiyaki consists of several software elements.

- Vine Linux 2.1CR (Red Hat 7.2 compatible)
- Apache 1.3.x
 - mod_perl 1.2x
- Perl 5.6.x
 - Jcode.pm (for Japanese)
- ImageMagick 5.3.x
 - FreeType 2.0.x
- Some True Type Fonts.





Merits of Mojiyaki

Speedy generation of images.

(implemented by mod_perl)

Implementation of any character code:

Unicode, various local codesets, and **ASCII transliterations** used by researchers.

Freely generation of glyph images that follow respective character codes.



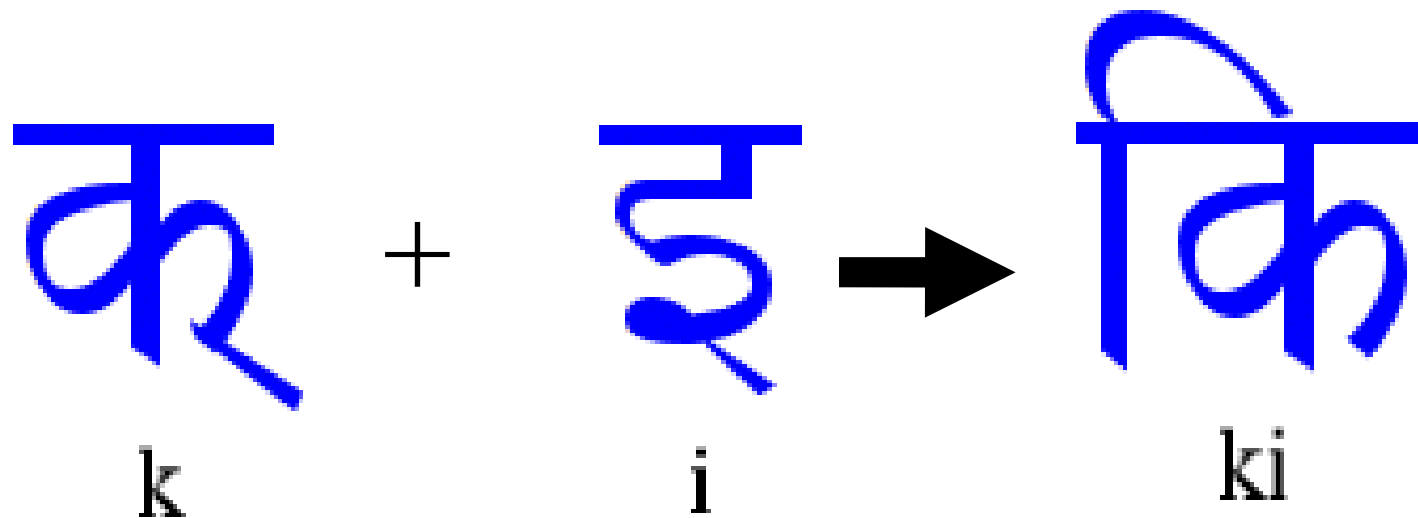
文字焼き
Multilingual Imager for Web

- ASCII transliterations:
 - Researchers, particularly Indologists, have created various kinds of data and analytical tools, transliterated into ASCII.
 - Writing order is only for left-to-right.
 - Useful for computer analysis.
 - However, most of Indic scripts involve complicated ordering, called “ hindi reverse” , and many ligatures.





ASCII transliteration and glyph images (1)



“Hindi reverse”





ASCII transliteration and glyph images (2)

क

k

ष

\$a



क्ष

k\$a

“Devanagari Ligature”



文字焼き

Multilingual Imager for Web

- Mojiyaki is able to generate all ligatures, by applying appropriate rules.
- It is able to modify characters, with ImageMagick features.
- Utilizing ASCII transliteration, it is possible to display any script on a Web page.
- CJK scripts can be displayed by URI-encoding (=ASCII data).





A HTML Tag to represent “Tibetan Script” in Tibetan .

```

```

བོད་ཡིག་

- Some parameters are available to control size, color, or image types.
- This ASCII transliteration (bod yig) complies with the “extended Wylie” rule.



文字焼き
Multilingual Imager for Web

Mojiyaki is useful for:

- displaying multi-script data on old clients.
- dynamic contents, such as electronic dictionaries.





Examples of Web sites utilizing Mojiyaki

- “**Tibetan**-Japanese Electronic Dictionary”
(Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies)
→ <http://www.aa.tufs.ac.jp/~hoshi/>
- “Welcome to **Hindi** World”
(also at ILCAA)
→ <http://www.aa.tufs.ac.jp/~kmach/>
- “**Tibetan-Sanskrit** Search to the Saddharmapu.n.darika (eDic)”
→ <http://suzuki.ypu.jp/edic/>



[eDic (pre0.4):: Search Results]

search results

[Search: bden pa]
Total 795 blocks match.

[1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32]
[33] [34] [35] [36] [37] [38] [39] [40]

[1] **2219 point.** (sdp.39824) (/1:1.09861228866811 2:-0 3:/1:3.17805383034795 2:-0 3:1)
[Skt:: c22-7-21=c22-7-22] []

येन सत्येन सत्य वचनेन स्वम् मम बाहुम् तथागत पूजा कर्मणे परित्यज्य सुवर्ण वर्णा मे कायो भविष्यति
तेन सत्येन सत्य वचनेन अयम् मम बाहुर यथा पौराणो भवतु

(yena satyena satya-vacanena svam mama baahum tathaagata puujaa karma.ne parityajya suvar.na var.no me kaayo bhavi.syati tena
satyena satya-vacanena ayam mama baahur yathaa=pauraa.no bhavatu)

[Tib:: 243a5-243a6] [2 219.72]

བདེན་པ་ ཅང་ བདེན་པ་ རི་ཚིག་གང་གིས་ བདག་ གི་ལག་པ་ དེ་བཞིན་ གཤེགས་པ་ལ་མཚོན་པ་

རི་ལྷིང་ཡོངས་སུ་བདང་ བསམ། བདག་གི་ལུས་གསེར་གྱི་ཁ་དོག་དུའགྱུར་བ་ རི་ བདེན་པ་ ཅང་ བདེན་པ་ རི་ཚིག་དེས།

བདག་གི་ལག་པ་འདི་ སློན་གྱི་བཞིན་དུ་གྱུར་ ཅེག་

Conclusion

- “Mojiyaki” is an improvisation technique, which may shortly become obsolete, as better systems become available.
- However, although I made a similar prediction at a conference two years ago, I have just utilized this for Loas within the last three months.
- Thus, Mojiyaki may still be needed for some years to come.