

Writing Javanese Script in HTML using Unicode True Type Font and JawaTeX

Ema Utami

Magister of Informatics Engineering of STMIK AMIKOM Yogyakarta
Ring Road Utara ST, Condong Catur, Depok Sleman Yogyakarta Indonesia

ABSTRACT

Since Javanese script is added into Unicode Standard in version 5.2 at October 2009, there are opportunities to many applications using Javanese Unicode standard. Javanese typeface (font) also already developed according to Unicode standard. The problem using Javanese Unicode standard is not easy to write Javanese in computer media. Transliterator program named JawaTeX can perform transliterate from Latin to Javanese script. The JawaTeX program produces several documents: parsing document, TeX document and post script document as final document that can be port into pdf document. Web based JawaTeX also developed to produce image based Javanese script embedded on HTML. JawaTeX can be modified to produce HTML format based on Javanese Unicode. The HTML file is constructed by numeric character reference (NCR) associate to Javanese Unicode standard. Having Javanese Unicode HTML document is more portable and flexible. Using Unicode typeface and JawaTeX capability and NCR this paper will introduce a concept to build Javanese transliterator having HTML based file as an output.

General Terms

Pattern Matching, Javanese Unicode, TeX

Keywords

Javanese script, Unicode typeface, JawaTeX, transliterator, NCR

1. INTRODUCTION

Writing Javanese script is rarely done recently, difficulties in how to write Javanese script is one most reason. The digital era can treat this weakness, the computer media now day can be used to write Javanese script, several programs can be used such as Hancaraka, Pallawa or directly using Unicode typeface for Word Processor. Writing Javanese letter by using the font, the user must have good knowledge in reading and writing Javanese letters and the user is expected to memorize the symbols and certain complicated Latin letters to obtain the expected Javanese letters [7]. Writing Javanese script in computer media technology such as word processors still have difficulties, to produce a Javanese letter could be tricky. An example, to write “*jer basuki mawa bea*” using Unicode typeface, user must be type on the keyboard “[J] [Shift+Q] [Shift+Z] [B] [S] [Shift+U] [K] [Shift+I] [M] [W] [B] [Shift+E] [Y]”.

2. JAVANESE UNICODE STANDARD

The Unicode Standard is the universal character encoding standard for written characters and text. It defines a consistent way of encoding multilingual text that enables the exchange of text data internationally and creates the foundation for global software [4]. The Unicode Standard specifies a numeric value and a name for each of its characters, represented in one of three encoding forms: a 32-bit form

(UTF-32), a 16-bit form (UTF-16), and an 8-bit form (UTF-8). The Unicode Standard, Version 6.0, contains 109,384 characters from the world’s scripts [4], Javanese script was added to Unicode in October, 2009 with the release of version 5.2 [5]. The Javanese character block is defined by The Unicode Standard with the range from A980 to A9DF in Javanese Unicode Chart [3]. Part of Javanese Unicode chart is

	A98	A99	A9A	A9B	A9C	A9D
0						
1						
2						
3						
4						
5						
6						
7						
8						
9						

presented on Fig 1.

Fig 1: Part of Javanese Unicode Chart

A Unicode typeface is a computer font that is mapped into the Unicode standard Character Set. The main reason using of Unicode typeface is fact that the international scope of computing, digital information interchange, and electronic publishing has created a need for world-wide character encoding standards. Unicode is a comprehensive standard designed to meet such a need [2].

There are two Unicode base fonts for Javanese script, first font is named adjiksaka.ttf [1] and font created by R.S. Wihananto named Tuladha Jejeg Font [8]. Using those fonts, displaying Javanese script will be easier, office application such as Microsoft Word or OpenOffice Writer can use font directly by selecting font type (Fig 2).

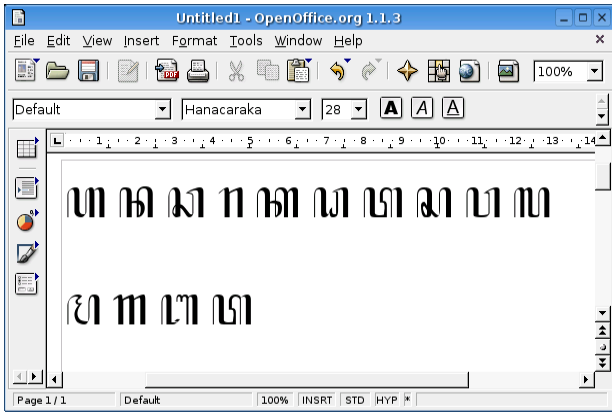


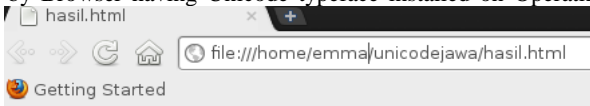
Fig 2: OpenOffice using Javanese Font

3. UNICOCE CHARACTER REFERENCE

Having Unicode typeface, HTML document can display Javanese by writing the code of Javanese Unicode using NCR. NCR is markup construct used in HTML that consists of a short sequence of characters, representing a single character from the Universal Character Set (UCS) of Unicode. The NCR written using two syntax "&#D;", where D is a decimal number, and "&#xH;" or "&#XH;", where H is a hexadecimal number [9]. Writing Javanese script in HTML code then must be using the NCR plus associate number in Unicode. Letter CA in Javanese Unicode chart occurs in U+A995, RA in U+A9AB and KA in U+A98F. HTML code for Javanese script "CA RA KA" letter is "ꦏ".

```
<HTML>
&#xa995 &#xa9ab &#xa98f
</HTML>
```

HTML code above will produce Javanese script while accessed by Browser having Unicode typeface installed on Operating



System (Figure 3).

Fig 3: Javanese script on web browser using Unicode typeface.

4. MODIFIED JAWATEX

The main problem in writing Javanese script on HTML is written the code. It is very hard to remember the code of Javanese letter at Unicode. JawaTeX is intended to write Javanese character simpler and easier based on complex transliteration algorithm. The schema process of Latin to Javanese character transliteration with LaTeX is in figure 4

[7]. The JawaTeX transliterator concept can be modified to produce HTML code containing Unicode Javanese letter. The intermediate text contains special TeX format resulted by JawaTeX program that can be modified using parsing program and result a HTML code. Modified JawaTeX and Unicode typeface can be a solution to write Javanese script easily.

Intermediate text is document with TeX/LaTeX code and syntax. The code and syntax will be used to transliterate split Latin string pattern that the Javanese character positions are known. Intermediate text is a text document with extension .tex that follows right rule to write Javanese character [6]. Example of intermediate text produce by JawaTeX is presented below,

```
\documentclass{article}
\usepackage{JawaTeX}
\usepackage[none]{hyphenat}
\begin{document}
\noindent
\raggedright
\begin{jw}
\ha\na\ca\ra\ka
\da\ta\sa\wa\la
\pa\dha\ja\ya\nya
\ma\ga\ba\tha\nga
\end{jw}
\end{document}
```

Structure of intermediate text can be divided into 3 parts: Header, Code and Footer (Fig. 5). Header on intermediate text consist preamble of LaTeX document followed by "begin document" statement and beginning JawaTeX code. Footer consists of ending of JawaTeX and ending of LaTeX Code.

5. RESULT

By developing a program, the intermediate text can be modified from TeX document to HTML document. The HTML based file will have Javanese Unicode typeface. Header and Footer on intermediate text can be replaced by HTML based header and footer. The structure of Javanese Unicode HTML file will be:

```
<HTML HEADER>
<NCR JAVANESE UNICODE>
<HTML FOOTER>
```

Perl based program was developed to convert intermediate text to HTML document. Intermediate text is read by Perl program and removes the Header, Footer and replace by header and footer HTML. The flowchart of conversion process is presented on Fig 6. After Header and Footer were removed from intermediate text, rest of file is the JawaTeX code. JawaTeX code is code to produce correct Javanese document using JawaTeX algorithm. Process of conversion is only replacing JawaTeX code by Javanese Unicode.

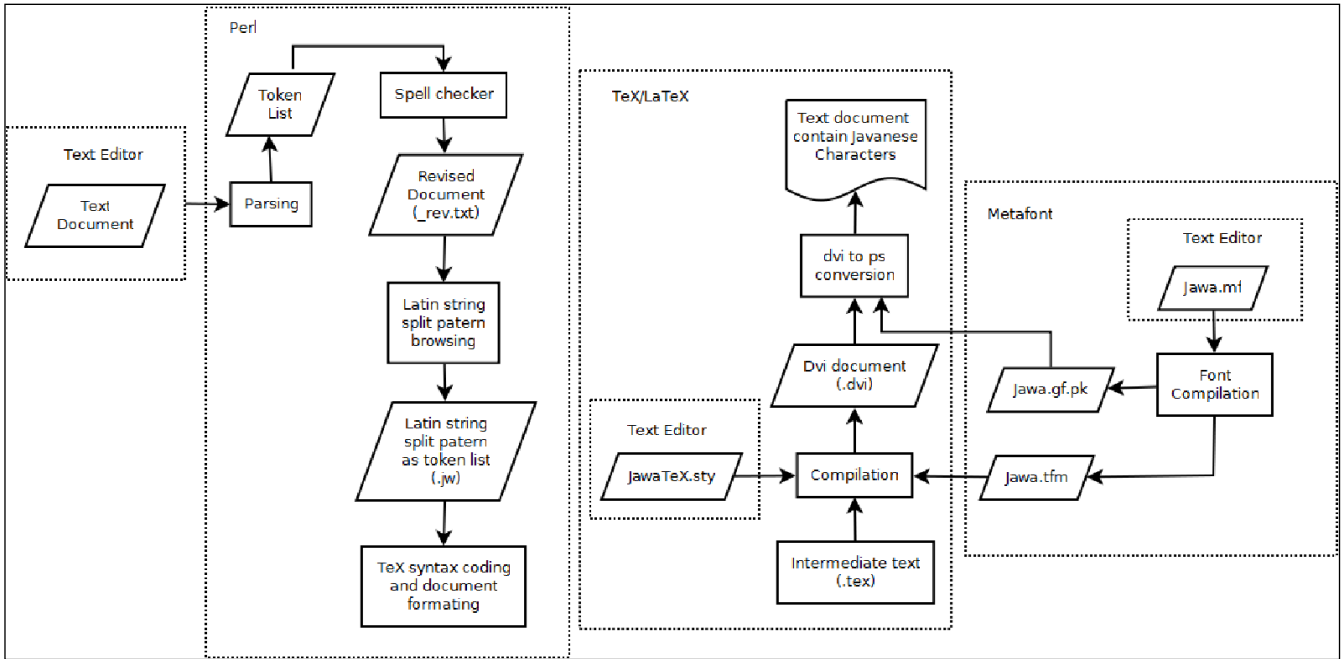


Figure 4: The schema process of Latin to Javanese character transliteration with LaTeX

Javanese Unicode and LaTeX code are writing on data file call Data Unicode. Data Unicode contain pair of key and value, key is JawaTeX code, and value is Javanese Unicode represented by NCR.

```
\sa : &#43441;
\wa : &#43438;
\la : &#43437;
\pa : &#43429;
\dha : &#43427;
```

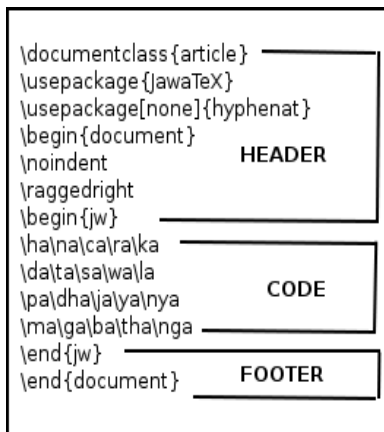


Fig 5: Structure of Intermediate Text

By using hash of array from Perl programming, program operates Pattern Matching of two arrays. If code in intermediate text is found on key in data then value on data file is replace the code. The process repeat until end of JawaTeX code. Data Unicode file is pair of key and value separated by “:” sign, example of data is:

```
\na : &#43428;
\ca : &#43413;
\ra : &#43435;
\ka : &#43407;
\da : &#43426;
\ta : &#43424;
```

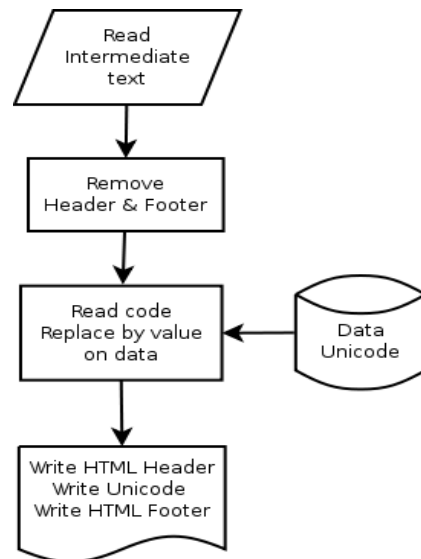


Fig 6: Flowchart Conversion Process

Example of file source.txt, a intermediate text file produced by JawaTeX program that will be processed by perl program is :

```
\documentclass{article}
\usepackage{JawaTeX}
\usepackage[none]{hyphenat}
```

```
\begin{document}  
\noindent  
\raggedright  
\begin{jaw}  
\ha\na\ca\ra\ka  
\da\ta\sa\wa\la  
\pa\dha\ja\ya\nya  
\ma\ga\ba\ta\nga  
\1\2\3\4\5\6\7\8\9\0  
\msa\mna\mka\mnya  
\A\I\U\E\O  
\hu\he\hE\hi\ho  
\hrang\hyang\hrar  
\da\ta\pk\sa\wa\la  
\end{jaw}  
\end{document}
```

Perl program named it `j2html.pl` is developed to do conversion. The program reads `source.txt` and compares it with data file; the result of conversion is `target.html`. The `target.html` is a HTML file contain of NCR represent of Javanese Unicode standard :

```
<HEAD><META HTTP-EQUIV='CONTENT-TYPE'  
CONTENT='text/html; charset=utf-  
8'></HEAD><HTML><br>&#43442;   &#43428;  
&#43413;   &#43435;   &#43407;<br>&#43426;  
&#43424;   &#43441;   &#43438;  
&#43437;<br>&#43429;   &#43427;   &#43415;  
&#43434;   &#43418;<br>&#43433;   &#43410;  
&#43431;   &#43424;   &#43412;<br>&#43473;  
&#43474;   &#43475;   &#43476;   &#43477;  
&#43478;   &#43479;   &#43480;   &#43481;  
&#43472;<br>&#43439;   &#43423;   &#43409;  
&#43416;<br>&#43396;   &#43397;   &#43400;  
&#43404; &#43406;<br></HTML>
```

The running of program and result file is display on Fig 7.

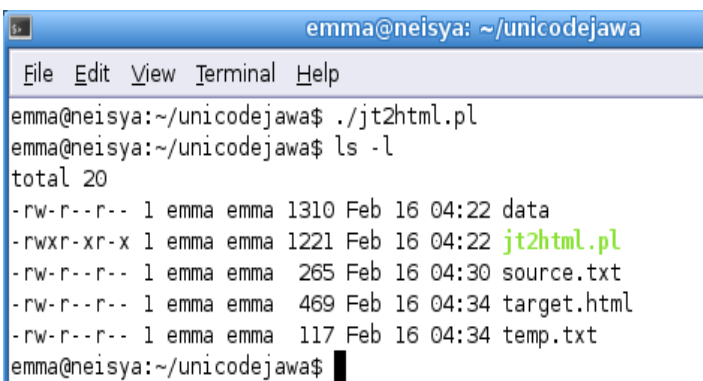
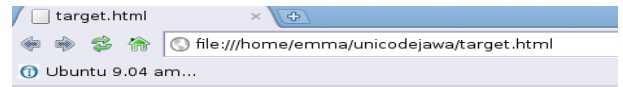


Fig 7: Running program and result files.

Result file, `target.html` can display Javanese script using a browser supported Unicode encoding and Javanese Unicode typeface installed on the operating system. Fig.8 is an example display of file `target.html` viewed by using Chromium web browser 6.0.472.63 (59945) Built on Debian 6.0.2, running on Debian 6.0.4.

If browser not support Unicode encoding nor installed Javanese typeface then browser can not display Javanese

script. Some browser version having supported Unicode encoding also can not display Javanese script correctly due to development of browser and typeface problems.



ꦭꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁ
ꦭꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁ
ꦭꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁ
ꦭꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁ
ꦭꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁ
ꦬꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁ
ꦬꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁꦫꦏ꧀ꦲꦸꦁ

Fig 8: Display result using Chromium web browser

HTML based Javanese script using Javanese Unicode can easily port into another document that supports Unicode encoding such as words processor: LibreOffice, Microsoft Office or OpenOffice. Result of Javanese script display on `fiFig. 8` can be easily copied and paste into LibreOffice Word Processor (`Fig. 9`). The result of copy will display exactly the same with source document, this is an advantage of using Unicode standard, document will be easily to port to another application.

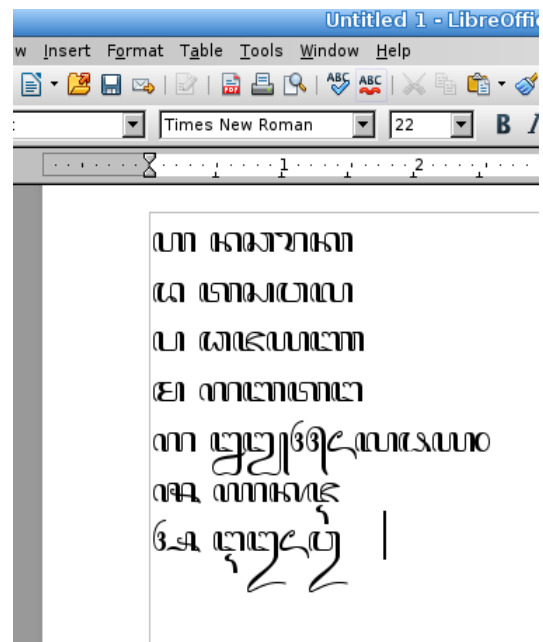


Figure 9. Copy Paste Result from `target.html` to LibreOffice

6. CONCLUSION

Intermediate text result by JawaTeX program can be converted to Javanese Unicode based on HTML format file using Perl program based on Pattern Matching algorithm. Structures of Intermediate text make it easier for program to convert JawaTeX code into HTML code.

Result of program a document based on Javanese Unicode HTML made document is more portable and flexible. Document will be easily to copy to another application as long as supported Unicode encoding format, for an example HTML document format to Word Processor document.

REFERENCES

- [1] Adjikaka Font, <http://www.adjisaka.com/download/adjiksaka.ttf> accessed Jan 26, 2012
- [2] Bigelow, C; Holmes K., *The design of a Unicode font*, Electronic Publishing, Volume 6, 1993
- [3] Javanese Unicode Characters code
<http://www.unicode.org/charts/PDF/UA980.pdf>.
- [4] The Unicode Consortium. *The Unicode standard, version 6.0.0*. Technical report, Mountain View, CA, USA, 2011.
- [5] The Unicode Consortium. The Unicode Standard, Version 5.2.0, defined by: *The Unicode Standard, Version 5.2*(Mountain View, CA: The Unicode Consortium, 2009. ISBN 978-1-936213-00-9).
- [6] Utami, E.; Istiyanto, J.; Hartati, S.; Marsono; Ashari, A., *JawaTeX: A system for Typesetting Javanese*, International Journal of Computer Science and Information Security (IJCSIS), ISSN: 1947-55000, Vol. 7 No.3 2010
- [7] Utami, E.; Istiyanto, J.; Hartati, S.; Marsono; Ashari, A., 2010, *Framework of JawaTeX*, International Journal of Computer Science and Network Security (IJCSNS), ISSN : 1738-7906, Volume 7, Nomor 4 2010
- [8] Wihananto, R.S, *Fonta Unicode Aksara Jawa, Implementasi Diterimanya Aksara Jawa ke Unicode Standard, Versi 1.11*, <http://sites.google.com/site/jawaunicode/main-page> accessed Jan 26, 2012
- [9] W3C, *HTML 5 A vocabulary and associated APIs for HTML and XHTML*, W3C Working Draft 25 May 2011 <http://www.w3.org/TR/html5/>