

# JawaTeX Web: Web Based Latin to Javanese Characters Transliteration System

Ema Utami

*Information System Major of STMIK AMIKOM Yogyakarta  
Ring Road Utara ST, Condong Catur, Depok Sleman Yogyakarta Indonesia  
Telp. (0274) 884201-884206, Faks. (0274) 884208  
[emma@nrar.net](mailto:emma@nrar.net)*

## Abstraction

Transliteration is a substitution letter by letter from one alphabet to another, free from how to actually speak those characters or it can be called a letter substitution or transliteration. JawaTeX can run in both Linux or Windows by make minor change to source code. This transliteration model can add a number of TeX based transliterator. By developing JawaTeX class or style in TeX, then the Javanese characters are expected to be equal with other ethnic characters such as ArabTeX, ChinaTeX, dan ThaiTeX and are most likely to be recognized by the global community.

JawaTeX is developed using several methods, The Context Free Recursive Descent Parser for parsing, rule-based for pattern splitting and Pattern Matching method for mapping into LaTeX class. This paper explain how JawaTeX work using web interface, the address of the site of JawaTeX can be found at <http://jawatex.org>. The results of Web based Latin to Javanese characters transliteration system are pdf document or figure that embedded on HTML.

The web transliteration provide 3 mode of transliteration: automatic mode, manual mode and embedded HTML. Automatic mode is done by uploading Latin text document that will be transliterated into pdf as result. The manual mode done by writing the input text on the HTML form, the result is also pdf document. The embedded HTML mode will transliterated the tex written in specific input format into figure on png format and inserted to HTML syntax.

**Key words:** web based transliteration, Javanese characters, embedded HTML, LaTeX, JawaTeX

## 1. Introduction

Javanese characters (script) is the second script used by approximately 71 million citizens of Yogyakarta, Central Java and East Java. But there was only very little effort to preserve it in the form of computer text

(digital). In many other countries, the research has been done to develop character processing for their local culture. This paper is influenced by the research of Free Open Source Software Localization (FOSS) [3] to develop software based on where the software is built. According to The Localization Industry Standards Association (LISA), Localization encloses product building that is appropriate to target culture (region and language) where the products are sold [2]. ArabTeX is a system for computer based typesetting of texts in the Roman script, which may contain insertions in some right-to-left script as Arabic and Hebrew [6]. TeX/LaTeX based transliteration is a transliteration using TeX/LaTeX. Many researchers do some researches in TeX/LaTeX based on transliteration; such as ArabTeX by Klaus Lagally, ChinaTeX by Shujun Li and ThaiTeX by Manop Wongsaisuwan [11]. Metafont program is used to develop fonts that are used by ArabTeX [5].

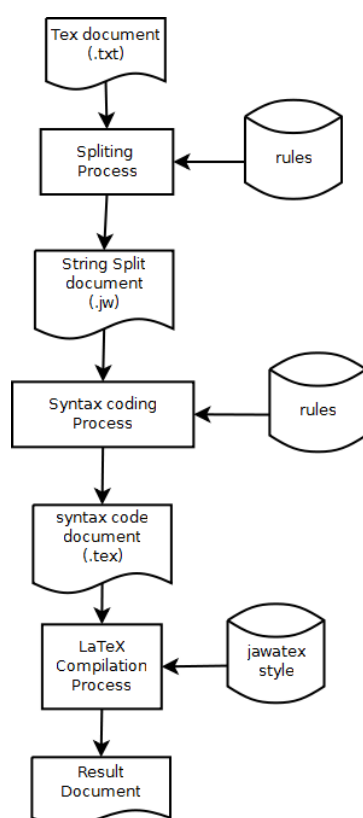
Until this research is written, there are three Latin-Javanese conversion program, namely Pallawa v 1.0, Hanacaraka v 1.0 and Carakan v 1.0.1. Several weaknesses from those three programs are the conversion result is not easy to be transliterated to other media or printed, not all programs can convert a text file and the result of conversion is not in accordance with the rule. In addition, not all Latin character writing can be transliterated to Javanese characters. There of them cannot written Javanese and Latin characters side by side and also can only be run in Microsoft Windows operating system.

## 2. JawaTeX

The development of JawaTeX is intend to build transliteration model of Latin to Javanese hopefully can transliterate all possibility of caharcter variation on input document. inputc characters variation. The methods using in this transliteration are, The Context Free Recursive Descent, rule-based and Pattern Matching.

The Context of Free Recursive Descent Parser algorithm is used to browse and split the text document [5]. A Rule-based method is used to develop Latin string split pattern list which result

from Latin text document processing. The Rule-based method is used to build several rules to handle the problems that can not be handled on previous researches. The Pattern Matching method is used to match each Latin string split pattern into LaTeX mapping format forms. The rule in transliteration model is made according to linguistic knowledge from a book guidance of writing Javanese script written by Darusuprata and published by Pustaka Nusantara Yogyakarta Foundation that cooperate with the government of the special region province of Yogyakarta, the government of central Java, and government of West Java [1]. The Rule-based method is used to build several rules to handle the problems that can not be handled on previous researches. The schema process of Latin to Javanese character transliteration with rule based is in figure 1.



**Figure 1:** The schema process of Latin to Javanese character transliteration with rule based

The Latin text document is parsed to determine the list of Latin string split patterns as token. The parser method used in this research is The Context Free Recursive Descent Parser. The Latin text document processing becomes the list of the Latin string split pattern by using rule-based method, whereas the matching process of each Latin string split pattern in

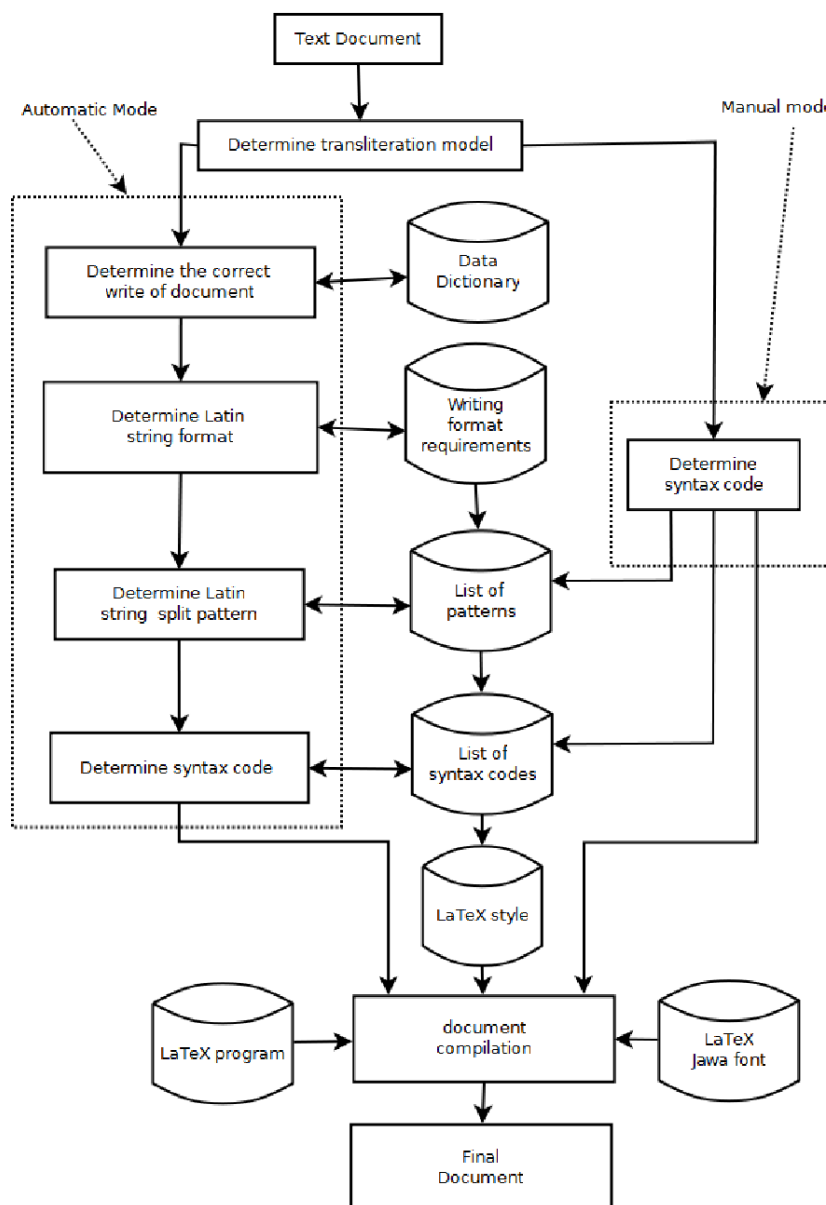
mapping form of LaTeX uses Pattern Matching method. With the rule-based method, the unsolved problems of the previous researches can be overcome by using certain methods. The established transliteration model is supported by the production rule of browsing the Latin string split pattern, the models of the Latin string split pattern, the production rule for the Latin-Javanese character mapping, the models of syntax coding pattern, style or macro LaTeX, Javanese character Metafont, and JawaTeX program consisting of parsing program and LaTeX style used to code LaTeX syntax. JawaTeX program consists of checking program and Latin string split to browse the Latin string pattern and LaTeX style which are used to code LaTeX syntax. Program transliterasi ini dapat dijalankan pada sistem operasi yang mendukung LaTeX dan Perl. Selain itu program ini juga dimodifikasi untuk dapat dijalankan melalui media web. The framework of Latin to Javanese character transliteration with LaTeX is in figure 2 [7].

Processing the Latin text documents into the TeX-based Javanese characters has two modes of transliteration, namely: automatic mode and manual mode. Automatic mode designed for users who do not have the knowledge to determine Latin string split and syntax coding patterns. Most of the stages of the process performed by the system, include [7]:

1. Determining the correct writing of Latin string in the source text by matching the source text and dictionary. This stage will result the Latin text documents that the writing of Latin string has been corrected.
2. Determining the formatting of Latin string by read, examining, modifying, altering, converting, inserting or adding other characters to meet the requirements of writing formats. The modification is only in the writing format form and does not change the meaning.
3. Determining the split string Latin pattern refer to 177 split pattern models that will produce 280490 Latin string patterns. This stage produce a list of Latin string split patterns that compose text document. The list of Latin string split pattern that has been obtained and then determined in the pattern of the relevant mapping transliteration to replace any Latin string split pattern into Javanese.
4. Determining the pattern syntax code which refers to the 57 coding syntax models. At the stage of correct pattern mapping, the first is to determine the position of Javanese characters as

the scheme of Javanese characters writing. Every split of the Latin string pattern can site the alphabet blocks consisting of 5 rows and  $n$  columns [6]. This stage produces a list of syntax codes that will be used for transliterated split of

the Latin string pattern which pattern layout has been obtained using the TeX/LaTeX format, are called the intermediate text.



**Figure 2:** Framework of Latin to Javanese character transliteration with LaTeX

After all 4 steps are performed automatically, the next stage is to compile the document. Intermediate text that has been obtained is compiled then *JawaTeX.sty* and *Jawa.tfm* are used by TeX to compile the document. *JawaTeX.sty* contains a Javanese script writing rules in a style TeX form, which includes [7]:

1. The word mastery which is different for example in a name.
2. The rule to combine the characters merger and define how to place and combine the characters.
3. Determining the shape of characters that is required in the merger because the Javanese characters have a lot of variety. A character in Javanese will have a different shape if it is placed in different positions despite having the same sound. A character in Javanese can also be possible to be paired with some Java characters

depending on the surrounding characters and the placement is not always in the back of the previous characters, but sometimes it must be inserted between the previous character. In addition, there are some characters that should not be paired with other characters, so that should replace the previous character. *Jawa.tfm* is font codes known by TeX and is a result of Metafont compilation.

Manual mode is intended for users who have the knowledge to determine the Latin string split patterns and syntax coding. There are 3 stages, the first is the correct of source text writing, the second is the writing of the Latin format string and the third is the split of Latin string patterns in which all of these have

been in the mind of users, who then arrange in the intermediate text that is ready to be compiled.

### 3. Web Based JawaTeX

Web based JawaTeX ia web interface for JawaTex. The site build using CMS Drupal to provide the user interface. On this web based transliteration provide 3 mode of transliteration : otomatic, manual and embedded HTML.

The automatic mode is done by uploading text document that will be processed JawaTeX as running in text mode (console). Text document that will be transliterated to Javanese character written using text editor and save as .txt documen. The document show on figure 3 is example for document source that will be transliterated.

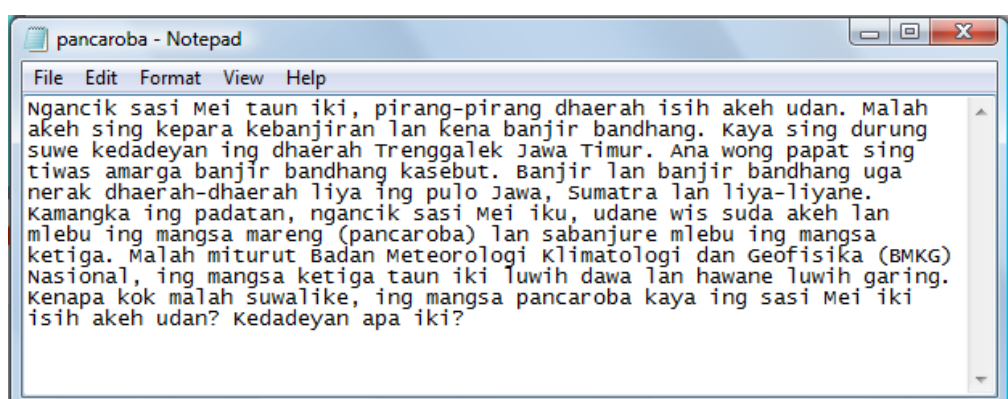


Figure 3: Text Document input

The process of transliteration is begin after text document is uploaded, The figure 4 show the upload input form. The process of transliteration on this stage same as show on figure 2 at automatic mode except determine correct word part. The web Jawatex on automatic mode not checking the correct words of document that being upload

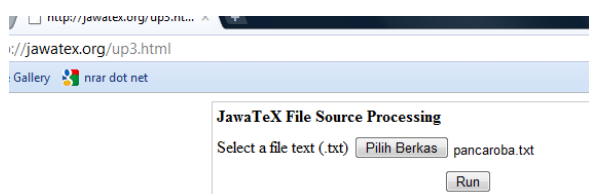


Figure 4. The Upload Form

The result of transliteration process is a pdf document. The link to the document is made to make document downloadable (figure 5).



Figure 5. Downloadable document link

The pdf document can be saved or viewed using pdf viewer such as Acrobat Reader. On this mode all text in input file will be transliterated into Javanese charaters as show on figure 6.

Manual mode on JawaTeX web is like automatic mode in transliteration process but the input process is differerent. On manual mode input is using HTML FORM instead uploading text file (figure 7).

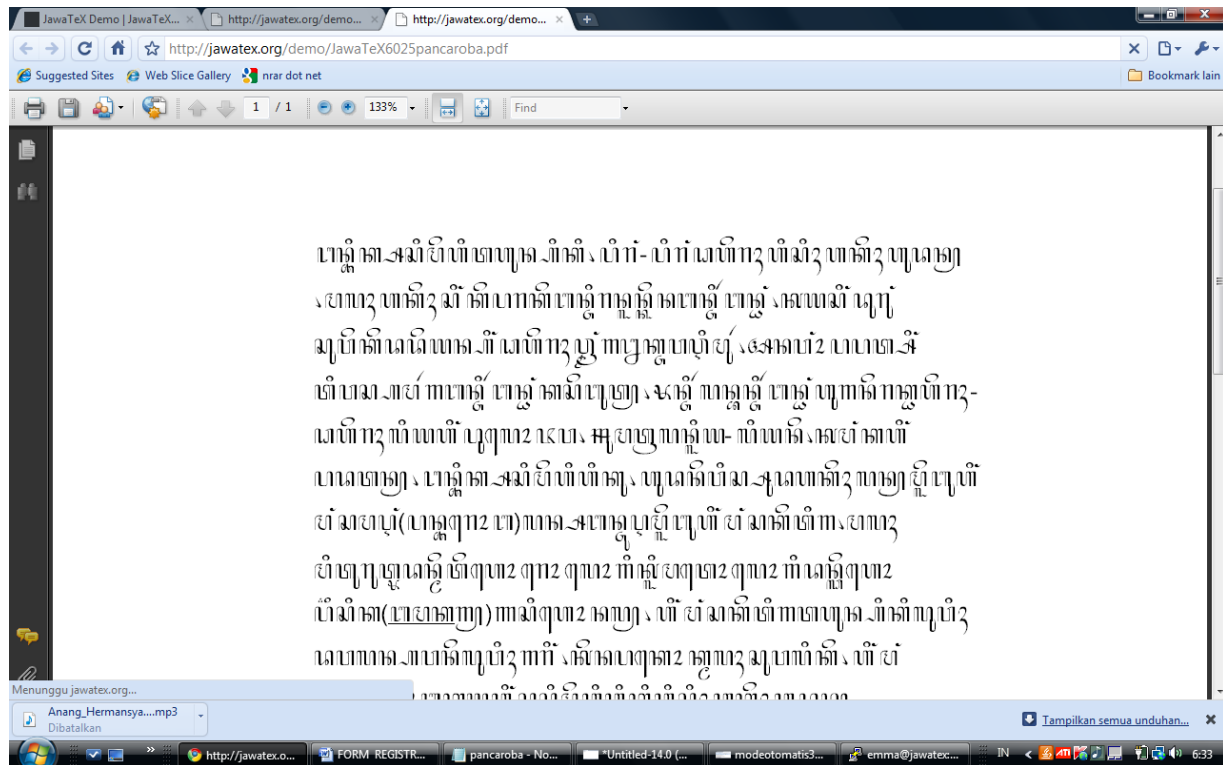


Figure 6. pdf document viewed

Process the form data to ...

http://jawatex.org/form.html

Suggested Sites Web Slice Gallery World Congress on ... Welcome to ACL 20... CONT

Laman ini dalam bahasa Inggris Apakah Anda ingin menerjemahkannya? Terjer

**Your Document Data:**

Enter your text here and use JavaTeX code such as  
`\jawa{ha\ka\sa\ra\ja\wa}` to write "Aksara jawa" in Javanese script

Type your text document above, Javanese code can be placed inside `\jawa{}` tags,  
 Characters inside tags turn into Javanese automatically after processing.  
 A complete howto at [HowTo link](#) ( still working on it)

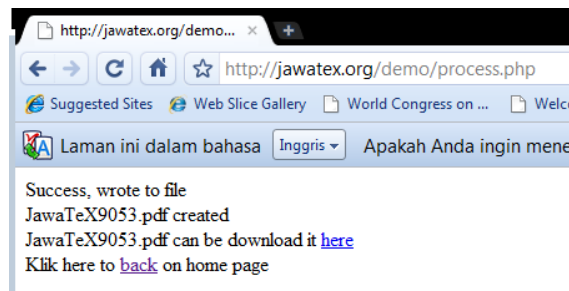
Kirim

Figure 7. HTML FORM input

This mode require user to know JavaTeX codes. JavaTeX codes is write using syntax `\jawa{codes}`. Only codes inside syntax JavaTeX will be transliterated into Javanese characters and rest of text

are not. After the form is submitted the text inside form will be saved and formatted as LaTeX document. The LaTeX document than processed into pdf

document. The link will be ready after pdf document created as show on figure 8.

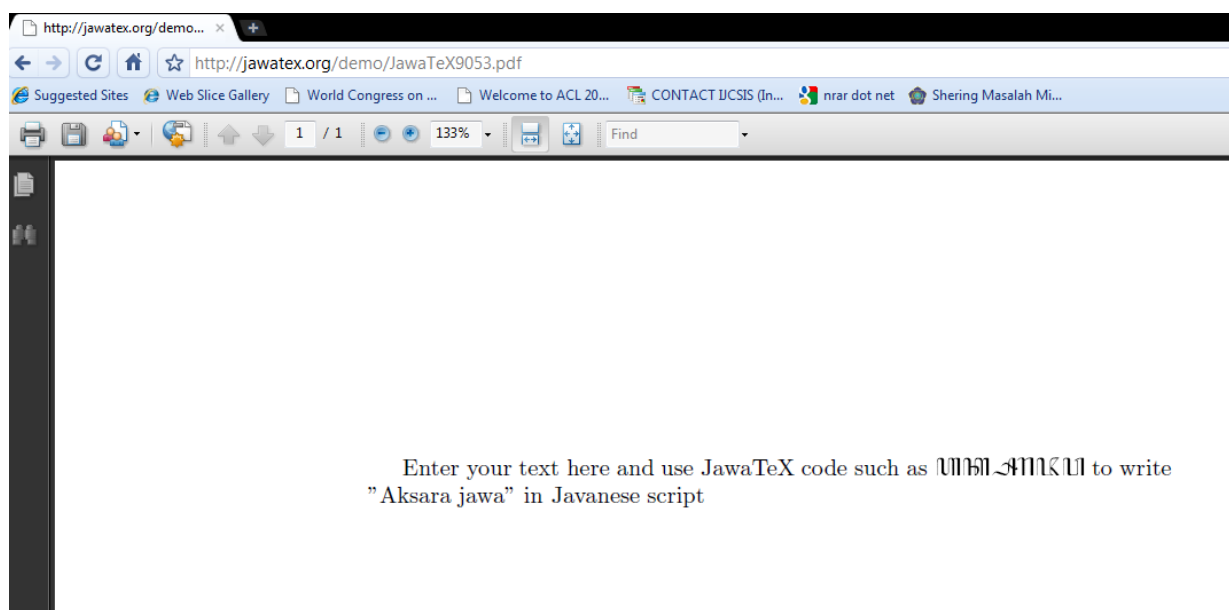


**Figure 8.** Pdf Document Link (manual mode)

The result of pdf document is differ form automatic mode. Link can be saved or viewed like in automatic mode. The manual mode will result pdf document contain Javanese characters and Latin characters. The figure 9 show pdf document that contain both Javanese and Latin characters.

#### 4. Embeded JawaTeX

In this mode using Drupal CMS and Drutex module. DruTeX is a powerful LaTeX module for Drupal. It can be used to display mathematics written in LaTeX as images inline with other text, or separately as a downloadable pdf.[8]



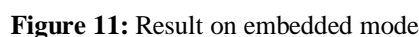
**Figure 9.** The manual mode result

JawaTeX web using this module to write Javanese characters inline with Latin characters in HTML format. Drutex module using `<tex>` and `</tex>` input format, the codes is place inside the input format.

The Drutex module is modified to use by JawaTeX. The codes JawaTeX are put in the input format. The work of modified Drutex module is show on figure 10. Every code in modified DruteX module will be process by JawaTeX to .dvi file and coverted into

image (.png). Each `<tex>` `</tex>` code will produce an image. Each image will have unique name for unique content. Every code inside `<tex>` `</tex>` will be saved as .tex file and hash to get a filename. The result, png file than inserted on HTML document. Example of this embedded HTML is show at Figure 11.





Education, museums, tourism, Heritage Protection Department, Institute of Traditional Culture Heritage which need transliteration from Latin document to Javanese characters for education, promotion, publication or publishing needed. This research result can be an educational tool and as step forward as well in the Java culture inheritance especially Javanese characters heritage in information technology era. The existing way to write Javanese characters is often inconsistent in character size and shape, by building the Latin to Javanese text document transliteration supported by the complete production rule to be able to handle the complexity in writing the Javanese

characters and can produce the good shape, beautiful and consistent Javanese characters.

The transliterator system framework include production rule of latin string split pattern browsing, string split pattern models, syntax code pattern models, LaTeX style, Javanese characters Metafont, and JawaTeX program package contain parsing and LaTeX style to write LaTeX syntax code. A JawaTeX program package contains two programs, checking and breaking Latin string to get the string split pattern and LaTeX style to write LaTeX syntax code.

The result of this research is able to perform the rules that haven't existed especially the consonant group findings, and finally the Javanese characters can be improved farther and completely. The implication ahead is that it is the time for Javanese writing from Latin spelling to be managed, the mechanism of using the program needs to be socialized well, the JawaTeX program resulted from this research gives an opportunity if it is adopted in Javanese linguistic. The concept of the text document split and the established transliteration in this article can be used as a basis to develop other cases. For the next research, the Javanese character split writing in good form still needs to be developed. The Javanese character writing sometimes cannot be justified alignment since the Javanese character writing does not recognize space between words.

The JawaTeX transliteration program can be run text mode (CLI) on Linux and Windows Operating System ( or all Operating System that support LaTeX and perl). This program also using web interface can be accessed at address <http://jawatex.org> to translate documents. Using modified Drutex module web based JawaTeX also show capability to write Javanese characters inline with Latin characters on HTML pages. Having web based interface make JawaTeX easier to use by everyone all around the world using Internet connection.

## References

1. Darusuprpta, et all., *Pedoman Penulisan Aksara Jawa*. Yayasan Pustaka Nusantara, Yogyakarta. 2002.
2. Lommel, A., *LISA The Localization Industry Primer Second Edition*. LISA-The Localization Industry Standards Association. 2003. Online at [www.lisa.org/interact/LISAprimer.pdf](http://www.lisa.org/interact/LISAprimer.pdf). 21 Februari 2006.
3. Souphavanh, A.; Karoonboonyanan, T., *Free/Open Source Software Localization*.

Elseiver, Reed Elseiver India Private Limited. 2005.

4. Utami, E.; Istiyanto, J.; Hartati, S.; Marsono; Ashari, A., 2008a, *JawaTeX: Javanese Typesetting and Transliteration Text Document*. Presented at Proceedings of International Conference on Advanced Computational Intelligence and Its Application (ICACIA) 2008, ISBN: 978-979-98352-5-3, 1 September 2008, page 149-153.
5. Utami, E.; Istiyanto, J.; Hartati, S.; Marsono; Ashari, A., 2008b, *Applying Natural Language Processing in Developing Split Pattern of Latin Character Text Document According to Linguistic Knowledge of Writing Javanese Script*. Presented at Proceedings of International Graduate Conference on Engineering and Science (IGCES) 2008, ISSN: 1823-3287, 23-24 December 2008, D5.
6. Utami, E.; Istiyanto, J.; Hartati, S.; Marsono; Ashari, A., *Developing Transliteration Pattern of Latin Character Text Document Algorithm Based on Linguistics Knowledge of Writing Javanese Script*. Proceeding of International Conference on Intrumentation, Communication, Information Technology and Biomedical Engineering (ICICI-BME) 2009, ISBN: 978-979-1344-67-8, IEEE: CFPO987H-CDR, 23-25 November 2009.
7. Utami, E.; Istiyanto, J.; Hartati, S.; Marsono; Ashari, A., *Framework of JawaTeX*, International Journal of Computer Science and Network Security, ISSN: 1738-7906, Vol. 7, No. 4, 2010.
8. Gutekunst D, *Drutex, a Drupal Module*, <http://drupal.org/project/drutex>.



**Ema Utami** received the S.Si, M.Kom and Doctoral degrees in Computer Science from Gadjah Mada University in 1997, 1999 and 2010 respectively. Since 1998 she a lecturer in STMIK AMIKOM Yogyakarta, Indonesia. Her areas interest are Natural Language Processing, Computer Algorithms, and Database Programming.