

C
001.
642
KOH

09

AUTOMATIC FONT DETECTION IN TAMIL DOCUMENT

A PROJECT REPORT PRESENTED BY

KENTHEESWARAN KOHILAN

to the Board of Study in Statistics & Computer Sciences of the

POSTGRADUATE INSTITUTE OF SCIENCE

in partial fulfillment of the requirement

for the award of the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

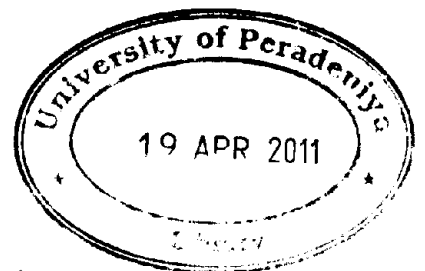
of the

UNIVERSITY OF PERADENIYA

SRI LANKA

2010

641582



AUTOMATIC FONT DETECTION IN TAMIL DOCUMENT

K.Kohilan

Advanced Technological Insitute

Varothayanagar

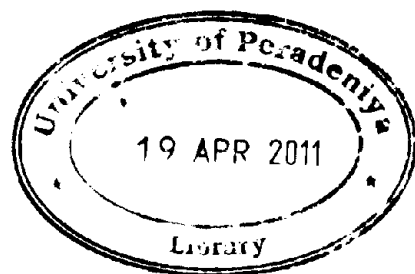
Trincomalee

Most of the people in the world use computers primarily for word processing. When we consider Tamil word processing, more than thousands of different type Tamil fonts have been utilized all over the world. All such fonts are not available in each computer at all times due to the familiarity or interest of fonts/font styles which might differ among users. All fonts do not consist of similar keyboard mapping due to different key assignment from font to font.

Suppose a person who prepared a Tamil text document using Microsoft Word in one computer, wanted to read it in another computer in different place. If the required font file is not available there, the document is in unreadable form in that computer. We can see similar type problems occur everywhere.

Normally this problem can be solved in two ways. One is by supplying the required font or set of fonts combined with the text document. Second is by identifying manually whether relevant or suitable fonts are available for substitution. If it is available on the machine, then it will automatically become readable. Otherwise manual effort should be needed to select suitable font. Both methods have drawbacks. Our objective is to provide a solution to solve this problem automatically without installing all Tamil fonts.

For this project, entire set of Tamil fonts were collected from many sources and analyzed manually to find out the keyboard mapping. Group classification of fonts was made based on the key mapping. In terms of identified font groups, font substitution can be made in the program.



The developed program can automatically convert a Tamil text document into readable form with little user interaction and minimum amount of time and storage. This program can be first installed in a computer and be used to read Tamil text documents.

Major limitation of this program is it can be worked in Rich Text Format (RTF). When text document is saved in other file extension than rtf, our program can not be used to read it directly. So, it should firstly be saved as .rtf file format in Microsoft word. Another limitation that is it does not support other graphical works.