

**M.Tech. (Computer Science) Dissertation Series 2005**

**DEVELOPMENT OF A WINDOWS-BASED BENGALI  
EDITOR WITH SPELL CHECKING FACILITY**

A dissertation submitted in partial fulfillment of the requirements  
for the M.Tech.(Computer Science) degree of the  
Indian Statistical Institute

*BY*

**SATADAL GHOSH**  
(CS0325)

Under the supervision of

**Dr. PALASH SARKAR**  
*Applied Statistics unit (ASU)*



**INDIAN STATISTICAL INSTITUTE**

203, Barrackpore Trunk Road  
Calcutta - 700035

## To Whom It May Concern

This is to certify that Mr. Satadal Ghosh of M.Tech. (Computer Science), second year student of Indian Statistical Institute, kolkata, has done his dissertation titled Development of a Windows based Bengali Editor with Spell Checking Facility under my guidance. This dissertation partially fulfills the requirement of M.Tech. (Computer Science) curriculum.

Date: 08.07.2005  
Place: Kolkata



(Dr. Palash Sarkar)  
Applied Statistics Unit  
Indian Statistical Institute (Kolkata)

# CONTENTS

<u>TOPICS</u>	<u>PAGE NO.</u>
CHAPTER- I : INTRODUCTION	1
CHAPTER- II : DESIGN OF WORK	2
CHAPTER- III : SOFTWARE SPECIFICATIONS	4
(A) USER MANUAL	3 - 9
(B) TECHNICAL DETAILS	10 - 11
CHAPTER- IV : SPELL-CHECKING ALGORITHM	12
CHAPTER- V : SOFTWARE AND TESTING	16
CHAPTER VI : FUTURE GOAL	17
IMPORTANT LINKS	18
ACKNOWLEDGEMENT	

---

## **INTRODUCTION**

Man has invented machines to facilitate his work and Computer is really a boon to the mankind. Man has created language to convey his thoughts to others. As population increased, printing became necessary to make more and more copies of books, letters, thesis etc to convey them to all. When computer comes, writing something and printing them becomes very easy.

Lot many editors are available for English language, which can check the spelling as well. But in Bengali, the number of good editors is not large and even the available ones are sometimes very costly. Further a few of them provide spell-checking facility in Bengali.

Here in this work , we have implemented a software which provide the user a good editor where he can write in Bengali letters and can have the option to check the spelling of his writing.

## **Design of Work**

Initially we studied some of the available softwares. Among them “BWEEdit” on Linux platform and “Lekho” need to be mentioned. Then we designed a user-friendly graphical user interface for the editor. Then we make use of Bengali true type fonts for the editor. A popular Keyboard map is then implemented within it. The general options for an editor like opening and saving a file, choosing fonts and size, making selected portions of writing bold, italics, underlined, printing the documents etc are incorporated gradually.

Next task was to implement the spell checking option. It is really a tough job because of the complexity of structure of Bengali language and variability of Bengali spelling. We made a suitable spell-checking algorithm and then implemented it to provide the editor a useful spell-checking facility.

### **Compatibility and Usage :**

Our software is compatible with the general DTP softwares. The files saved in RTF format can be opened and edited in the well-known PageMaker software. So, one can really arrange the document properly after writing it in our editor and having spell-checked.

This software runs in window platform. We have chosen Window platform so that our software can be actually used in publishing industry as most of the industries use Windows.

So, it is actually possible to use it in Bengali publishing industry. No known spell checker is available to the publishers. In that respect it may be possible for the publisher to use our software with spell checking option.

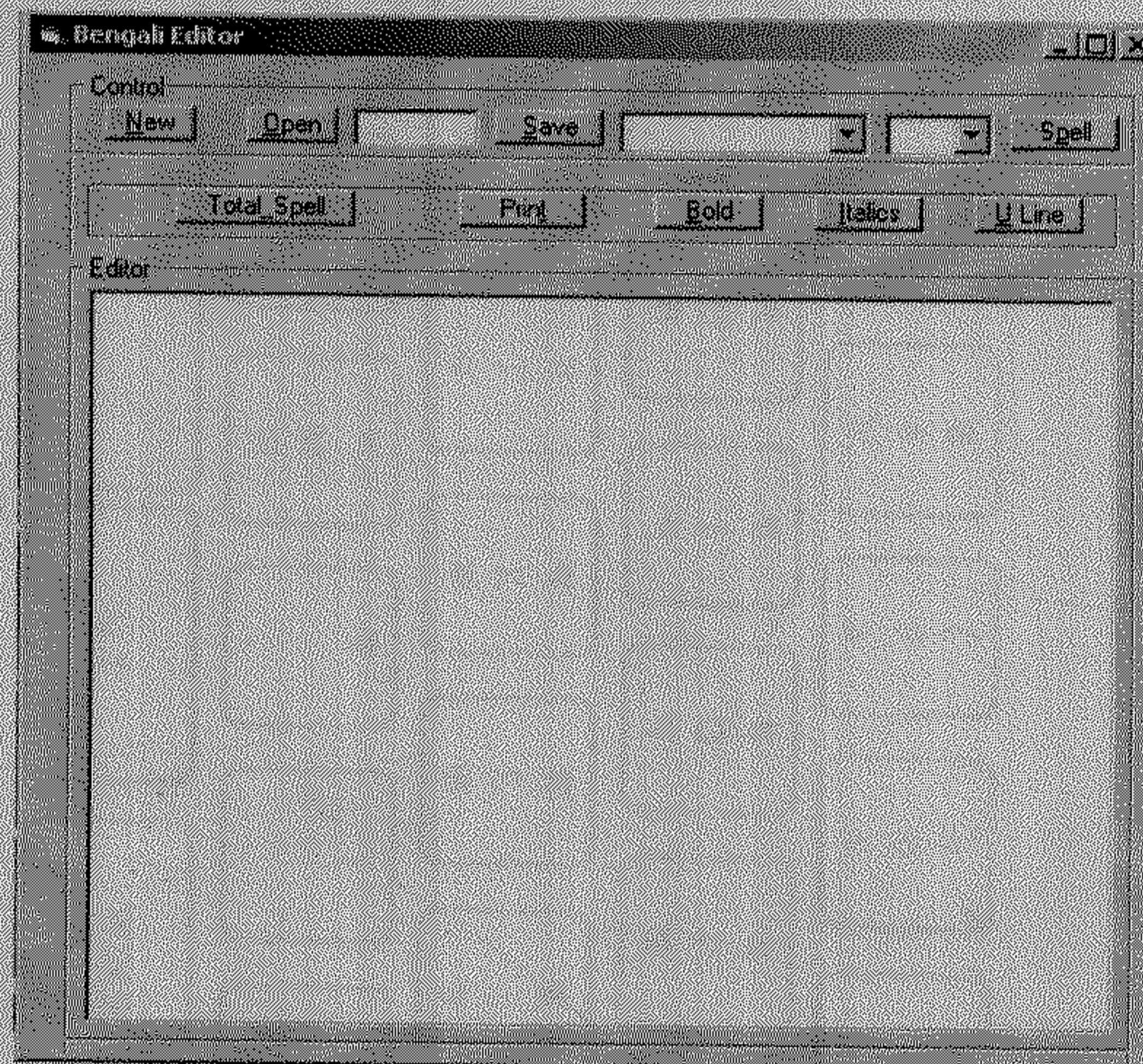
## Software Specification

### (A) User's Manual

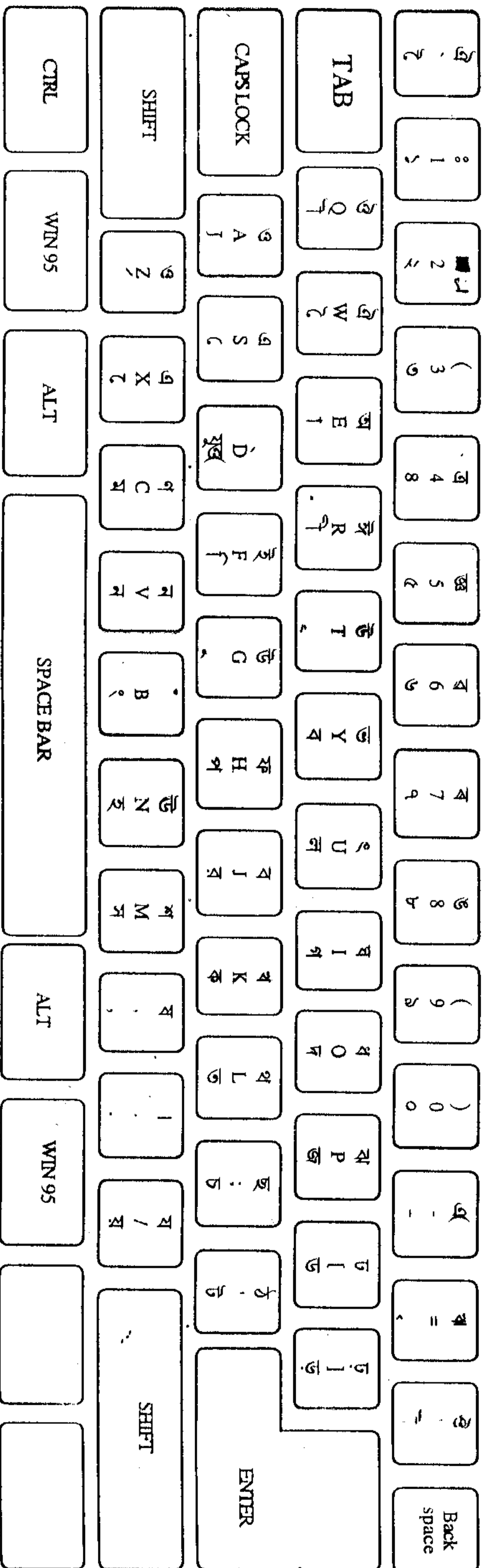
This software works in Window Environment. At first user has to install the true type Bengali fonts provided with this software. One can copy the fonts and simply paste it in the " Fonts" folder of Windows. Now the machine is ready to access those fonts.

On the second step, user has to click the exe file provided in this software to run it. The codes are also available with it. User has to install Visual Basic 6 first to see and modify the codes.

After opening the software user would get a GUI as shown below.



In the editor part, one can actually type Bengali directly from keyboard using the keyboard map given in the next page.

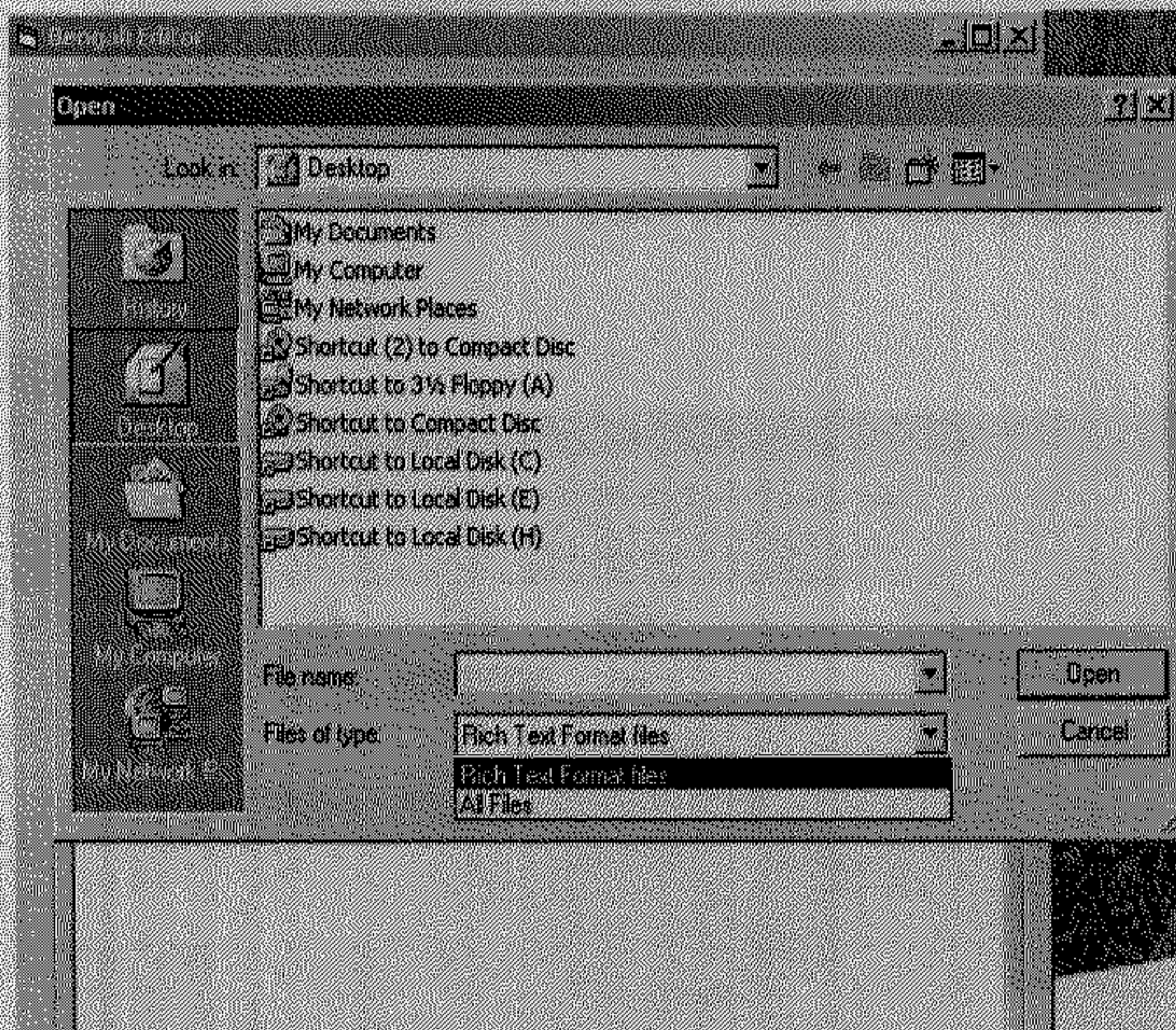




There are two combo boxes. User can choose his/her font for the editor from one combo box while the other one provides the user to select his/her own font size from 1 to 25.

On clicking "New" button, the editor is cleared for new writing.

On clicking "Open" button , a new window is opened as shown here,



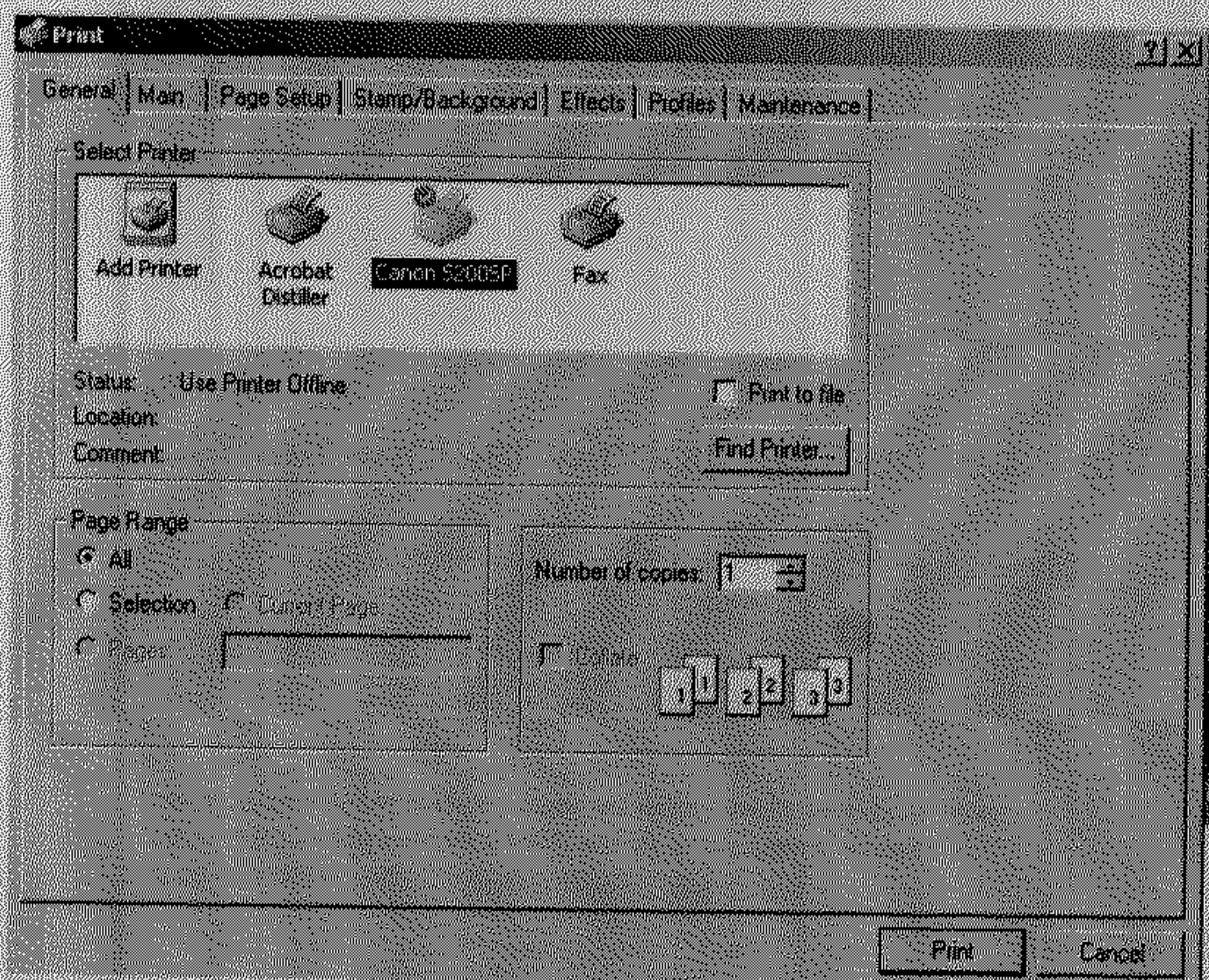
Here user can search and select his file which he wants to open in this editor. He/She has been given two options to open,

- All Files – All files are shown and file is opened in simple text format.
- Rich Text Format Files – Show all folders and RTF files, file is opened in RTF format, if feasible.

On clicking "Save" button, a similar kind of sub-window is opened where user can select the proper space where the file is to be stored by user given name and format. Here also, selection of "All Files" option saves the file in simple text format and selection of "Rich Text Format Files" saves it in RTF format.

User can make selected portion of text bold, italics and underlined by clicking respectively the "Bold", "Italics" and "U Line" buttons.

The "Print" button is given to print the document written in the editor. On clicking it another sub-window is opened as shown next.

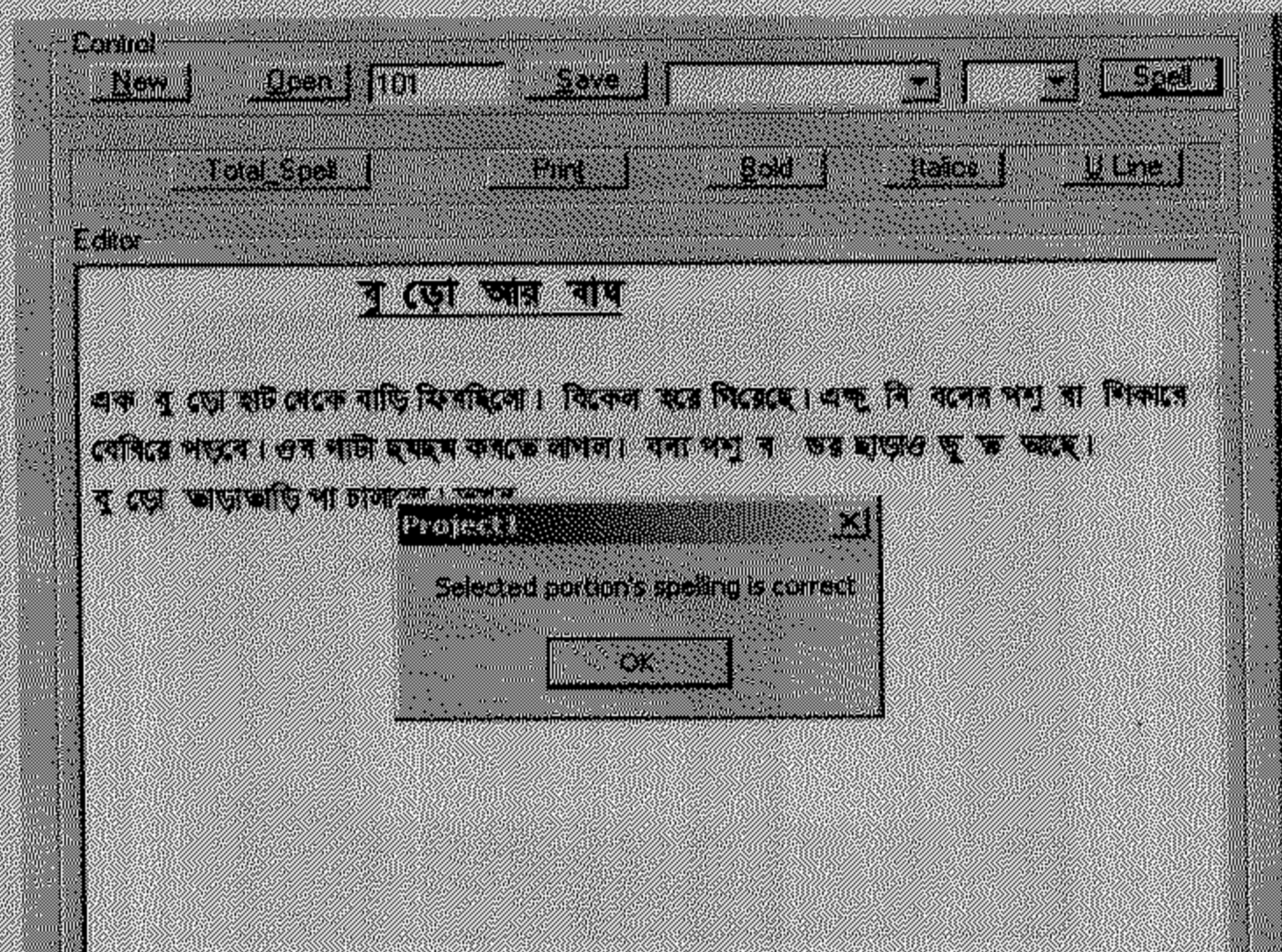


Here all the available printers are shown and one can select printer from the list to take the print of the document.

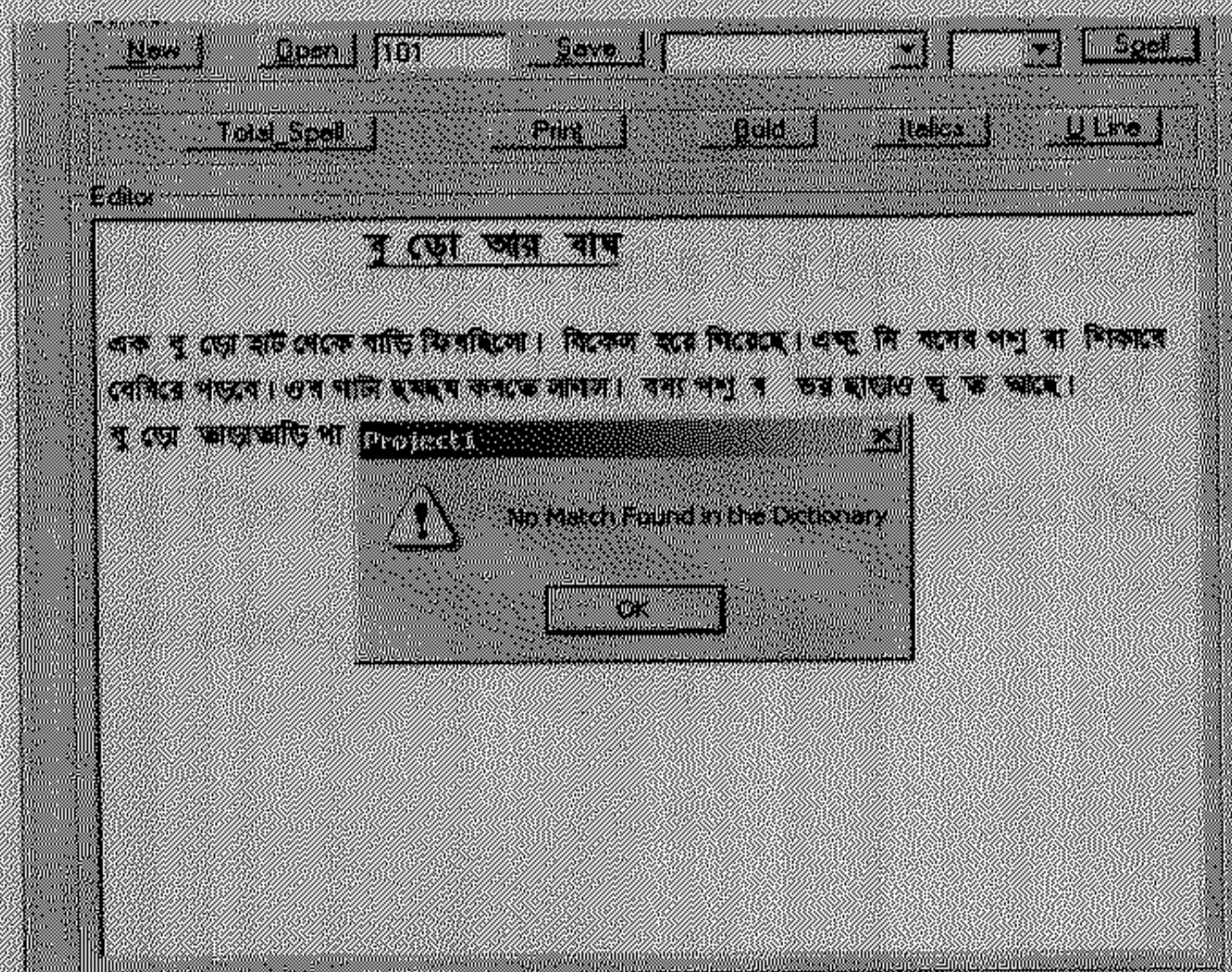
## Spell checking options:

User has been given two kind of spell checking options,

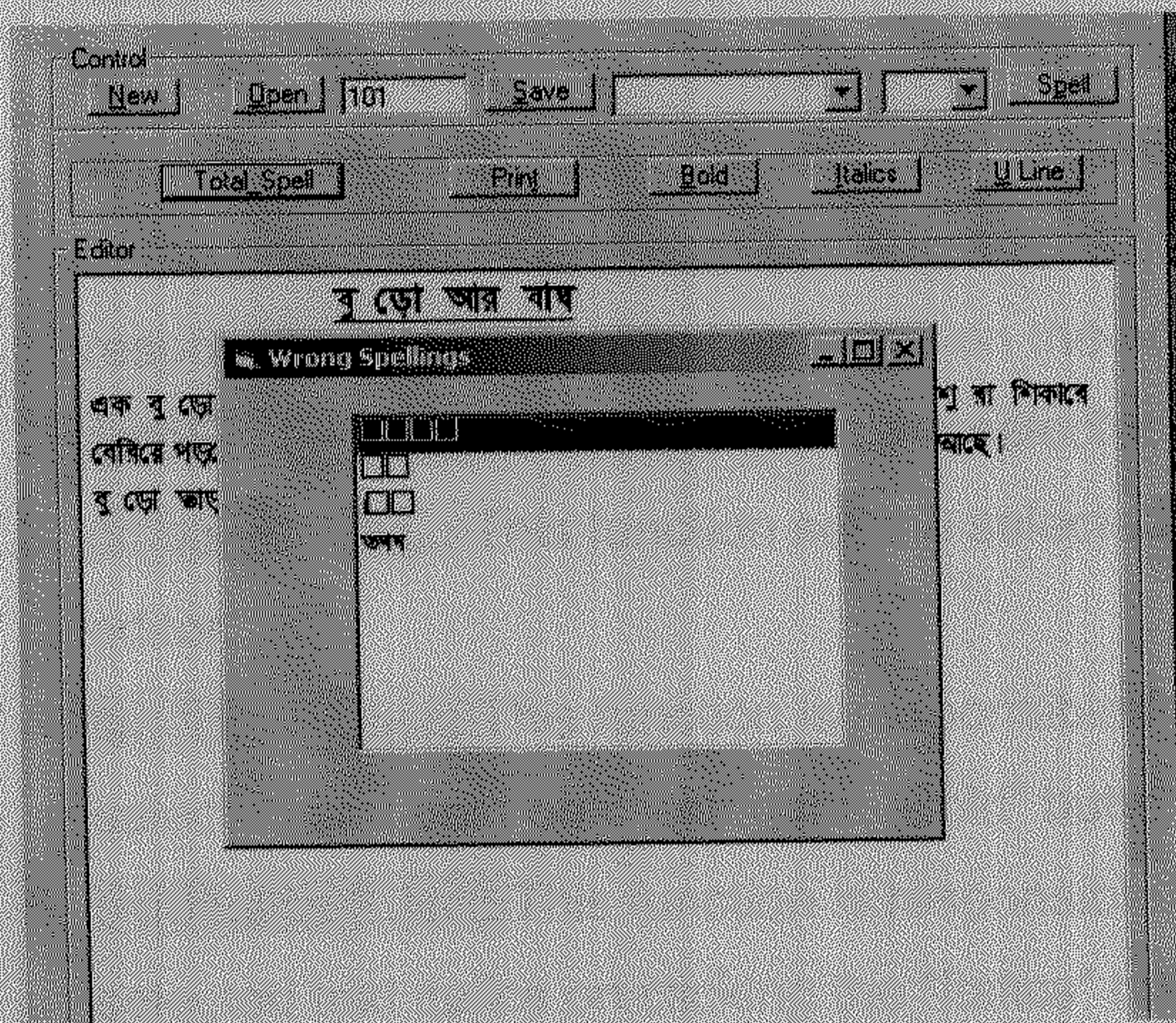
- The "Spell" button provides the user the option to check spelling of a selected word. If the spelling of the selected word is correct then the following message-window is opened.



If not then another window is opened as shown below.



- The "Total\_Spell" button gives the facility to check the spelling of the total document on the editor and after completion of the job, it gives a list of wrong spelled words. It would look as shown later,



It is to be noted that on first clicking the either of the spell button after opening the editor, it asks for the file storing the correct spellings of words. User is to select that file, spell.txt is provided with the software to serve the purpose. User can select his own spell-file for this purpose also.

## (B) TECHNICAL DETAILS :

The visual basic code of this software consists of different functions dedicated to different works. The function names, their work and interconnections of them are discussed here.

1. **Editor\_KeyPress**: This function plays a vital role in this software. It implements the keyboard mapping so that user can write in Bengali directly from keyboard. It takes the ASCII value of the pressed key and changes it to a suitable ASCII value responsible for the font to be shown on the editor.
2. **cmbFont\_Click**: This function set the selected font for the editor.
3. **cmbSize\_Click**: This function set the selected font-size for the editor.
4. **cmdBold\_Click**: This function is responsible for making the selected text bold on clicking the "Bold" button.
5. **cmdItalics\_Click**: This function is responsible for making the selected text italics on clicking the "Italics" button.
6. **cmdULine\_Click**: This function is responsible for making the selected text under-lined on clicking the "U Line" button.
7. **cmdNew\_Click**: This function clears the editor for writing new document on clicking the "New" button.
8. **cmdPrint\_Click**: This function works for showing the available printers and then printing the document with the selected printer after clicking the "Print" button.
9. **Form\_Load**: It is meant for initializing different variables whenever the software starts running.

10. **Form\_Resize**: This function always set the GUI position at the center of the window to make it visibly soothing.
11. **Open\_Click**: This function is responsible for opening a file in particular format whenever “Open” button is clicked.
12. **Save\_Click**: This function is responsible for saving a file in particular format whenever “Save” button is clicked.
13. **spell**: This is a vital function for this software. It actually implements the spell-checking algorithm, described later, within it. The other two functions described later actually use this module to perform their working.
14. **cmdSpell\_Click**: This function checks the spelling of a selected word after clicking the “Spell” button with the help of ‘Spell’ module.
15. **cmdtotSpell\_Click**: This function checks the spelling of the whole document at a time after clicking the “Spell” button with the help of ‘Spell’ module and at last shows a list of wrong-spelled or not available words.

## SPELL-CHECKING ALGORITHM

Here the spell-checking algorithm used in this software is discussed. We generally categorize the Bengali words into three categories, they are explained as follows:

Type 0 : Any word that is available in the dictionary/spell file is called type 0 word. Some examples are মানুষ , ধূর্ত , আমি , আমাকে , .....

Type 1 : Any noun available in the database appended by an allowable case endings gives a word of type 1. Examples are মানুষটি , মানুষটির , মানুষদিগকে .....

Type 2 : A word of type 2 is a concatenation ( as strings ) of two words. The first one must be a type 0 word and the second one must be a type 0 or type 1. Examples are চালডাল , ঘরবাড়ি.....

**Representation of words** True type fonts represent each word and checking is done with ASCII values of the fonts.

**The Spell File**: This software maintains a file called spell.txt , which consists of hundreds of words with correct spellings. This should be a copy of Bengali dictionary but presently lots of words are yet to be added in it.

**The Case Endings**: We are maintaining a list of valid case endings used in Bengali language in the file named caseending.txt. The list of those case endings is given below,

কে, খানা, খানাকে, খানাতে, খানায়, খানার, খানারে, খানি, খানিকে, খানিতে, খানির, খানিরে, গণ, গণকে, গণরে, গণে, গণেতে, গণের, টা, টাকে, টাতে, টায়, টার, টারে, টি, টিকে, টিতে, টির, টিরে, দিকে, দিককে, দিগের, দের, র, রা, রে .

Our spell-checking algorithm consists of four parts. These are stated as follows:

**Type 0 checking :** Read the whole word at a time and match it with the words given in the spell file. If a match is found then it flags TRUE and exit out of the algorithm. Otherwise go to the next part.

**Searching for longest match :** For each word written in the editor, our program read it character by character and search for a match at each stage. Whenever a match is found, the length information is kept in len\_max. If a longer match is found then len\_max is updated to store the longest match information.

**Type 1 checking :** After finding the longest match length, it tries to match the rest of the word with the allowable case ending list. If the remaining part matches with one of the valid case ending then it flags TRUE and exit out of the algorithm. Otherwise next part is executed.

**Type 2 checking :** Taking the longest match information, it tries to match the remaining portion of the input word with a valid word from the spell file. If a match is found then it flags TRUE and exit. Otherwise flag remains FALSE.



After exiting the algorithm, we say a word is correctly spelled if the flag is TRUE. Otherwise we say the spelling is wrong or not found in the dictionary.

We illustrate the above algorithm by some examples as follows:

Example 1: Take the word মানুষ . The first part of our algorithm take the whole word and finds a match in the spell file. As soon as it finds a match, it flags TRUE and exit. TRUE flag implies that it is a valid spelling. It is a type 0 word.

Example 2: Now take মানুষটি as an example. At first it checks for type0 word. When it fails, it finds the longest valid word which is মানুষ. Now the remaining portion টি is searched in the valid case ending list. As a match is found, flag becomes TRUE and the spelling becomes correct. It is a type 1 word.

Example 3 : Let us take the example চলডাল . First the type 0 checking fails. Then the longest valid word is found as চল . Now remaining portion ডাল is searched in valid case ending list and no match found. Then ডাল is searched in valid spelled words and a match is found. This makes the spelling correct. This is a type 2 word.

Note that the longest match becomes extremely important in this case as চা and চল are both valid words and we have to choose চল for declaring চলডাল as correct spelling.

Example 4 : Now for the word চলডালচালডাল , it is more than two concatenated words, so it fails to survive the type 2 testing, as here the remaining portion ডালচালডাল does not match with any valid word. So, the flag remains FALSE and spelling is not correct.

## SOFTWARE AND TESTING

Our Software contains the following parts :

- Project1 contains the main code of the software.
- caseending.txt contains 36 valid case endings used in Bengali language.
- spell\_1.txt contains more than two hundreds of valid Bengali words for checking of correct spelling.

### TESTING :

We have tested our software on writing down few pages of a book and running our spell-checking algorithm on it. In almost all the cases, our algorithm gives correct result. It has been observed that if the spelling list contains sufficient amount of words then this algorithm works well and good in almost all the cases.

## **FUTURE GOAL**

Our future goal is to incorporate the following features into our software,

- We would try to incorporate some grammar checking facility within it so that one can check grammar also.
- We would try next to give a number of nearly matching suggestive words in case of wrong or spelled words.
- A soft copy of dictionary database, if available, can be linked with it so that spell checking becomes better and one can get synonyms facility there. In that case a proper algorithm for searching dictionary database is an essential one due to huge number of words there.
- Our spell-checking algorithm can also be modified a bit for better performance in course of our further work.

## **IMPORTANT LINKS**

1. <http://tnp.saha.ernet.in/~pbpal/bangtex/bangtex.html>
2. <http://www.nongnu.org/freebangfont/donors.html>
3. <http://www.microsoft.com/typography/web/embedding/weft3/default.htm>
4. <http://www.microsoft.com/typography/developers/volt/default.htm>
5. <http://www.banglalive.com/mail/>
6. <http://lekho.sourceforge.net/>

## **ACKNOWLEDGEMENT**

I express my deep gratitude to Dr. Palash Sarkar , Applied Statistics Unit of Indian Statistical Institute, for his supervision, guidance, kind encouragement and all kinds of help in preparing this dissertation.

I am very grateful to Dr. D. P. Mukherjee of ECSU and Dr. Mandar Mitra of CVPRU of Indian Statistical Institute for their valuable guidance, suggestions and inspiration.

In this connection I would like to thank Dr. Abhijit Das, IIT Kharagpur , for valuable advice and inspiration at the initial stage of this work. I am very grateful to Dr. Arnab Chakraborty, Stat-Math Unit of Indian Statistical Institute, for his valuable advice and providing me necessary links and Bengali fonts in course of my work. I am thankful to my elder brother Anupam Ghosh for providing me keyboard map used in popular DTP softwares.

I would like to thank my friend Tilak Kr. Adhya for valuable as well as enjoyable long discussions with him in course of designing this work. I am also grateful to all the teachers of ISI from whom I have learned lot many things about computer science; all of my friends, with whom I have discussed about my work for their valuable thoughts and inspirations.

*Date : 07.07.2005*

*Satadal Ghosh*  
**SATADAL GHOSH**