Tibetan Typographical Specifications and Technical Realization Based on Word VBA

Hai-Lan DING\textsuperscript{a}, Kun-Yu Qi, Xiao-Long Zhao, Gui-Fu Xu

Northwest Minzu University, China National Institute of Information Technology, Lanzhou, Gansu, China

\textsuperscript{a}www.dinghailan@qq.com

**Keywords:** Tibetan Document Typesetting, Word VBA Template, Technological Realization.

**Abstract.** The key technology to realize the article unified layout lies in the format adjustment in the Tibetan document typesetting. At present, there is a phenomenon that the left and right borders cannot be aligned when the Tibetan version is typeset in the Word2010 office software. As a result, the Tibetan document plate does not conform to the traditional Tibetan writing norms and brings some limitations to office editing and ancient book collation. With Word software as the developing platform and Word VBA technology combined with VBA programming, the article realizes the proper arrangement and adjustment of the Tibetan documents in accordance with the writing format of the Tibetan version. It has great significance on the standardization of Tibetan document layout.

**Introduction**

In order to make the layout demand, use, convenience, all kinds of Tibetan documents, such as poems, Buddhist scriptures, essays, essay books, newspapers, periodicals and so on conforming to Tibetan writing standards are input in the Microsoft Word. According to the Word menu tools, the formatting are adjust. Due to the lack of right alignment in Tibetan typesetting, we can use Word VBA technology to program and achieve the standardization of Tibetan typesetting. Word VBA (Visual Basic for Application) is a programming language developed by Microsoft Corporation and attached to the office automation suite for general-purpose automation tasks [1]. The main role of VBA is to customize the functions and interactions in the application to help users to develop and exploit functional modules in the Office environment according to their own needs. Although Word provides powerful word processing and document control capabilities, the basic functionality of a Tibetan document that uses Word only involves a great deal of tedious typesetting under the environment where Himalayan IMEs are edited, and therefore an automatic form-factoring feature are required to amake the Tibetan version of the document standardize. Word VBA is an object-oriented, visual, event-driven, high-level language built on top of Visual Basic's framework. Its syntax, functionality, and even development environment are basically the same as Visual Basic. By Word VBA programming, Word documents can be automated. Through this method, the article studies on the compilation of Word VBA auxiliary standard and realizes the partial automation of standard writing.

**Application of Tibetan Writing Layout**

Many domestic research institutes and universities have developed and explored a number of word processing software since computer Tibetan system research and development began in the 1980s, which was a copy of Founder and Huaguang Tibetan DOS-based professional-type layout system. After 2000, with the popularization and application of Windows system, one after another word processing software appeared, such as Windows-based Tong Yuan, Ban zhi da, Qiong Mai, Sang bu zha. However they were neither international standard coding system nor compatible. Microsoft released a complex Windows-based text engine that made the Tibetan characters have a standard international encoding in 2007. The 2007 Word office software also fully supported the unified encoding of the Tibetan input system. Currently, the world-wide Microsoft Himalaya
Tibetan font is based on Unicode standard Tibetan fonts and any international Unicode fonts can be switched. At the same time, Tong yuan, Ban zhi da and other software and gradually withdraw from its application areas.

2. The Tibetan Layout Problem in Word

Tibetan text in the Word 2010 document used the Himalayan input method editing when a paragraph of Tibetan sentences selected decentralize paragraph format and the right can not be aligned. This phenomenon makes the Tibetan text in the Word2010 document Himalayan input method editing occurs when a paragraph of Tibetan sentences decentralized paragraph format selection occurs when the right can not be aligned. This phenomenon makes the articles of Tibetan scholar normalization and neatness in the editing of manuscripts, the publication of books, the writing of essays, the input of large quantities of manuscripts and the possession of Tibetan materials, so it is extremely important for typography of Tibetan documents. Due to the incompatibility of Tibetan typesetting, we can customize the functions in the application based on the programming language of Word VBA automation task so that the Tibetan documents can be typeset and achieved certain standardization.

Tibetan Typesetting

Typesetting is the process of making the layout work by adjusting the position size of the text, images, graphics and other visual information elements. Many informational manuscripts are required to be formatted according to a unified format when inputting a Word document, such as font size, font, paragraph, style, arrangement, page background, theme, and alignment. Sexual typography. Those manuscripts are mainly Tibetan newspapers, books, essays, and Buddhist scriptures, etc. The purpose of typesetting is to achieve uniformity and orderliness at the time of publication so that the format of the article can be read at a glance by readers.

Limitations of Tibetan Typesetting

The input method must be unified so the Microsoft Word software can achieve the standardization of the Tibetan version. Tibetan writing specifications of Tibetan various types of documents such as poetry, scriptures, essays, essay books and newspapers are input in the unified Word editor window. These special Tibetan symbols that apply to all kinds of style and their placement to the Tibetan typesetting, especially the style of the Book of Songs are trouble for Tibetan version of the standardization and documents alignment when the input document appears in the Word2010 office software layout process by the Himalayan input method editing.

Word Application Technology

It is not satisfied that complex document process problems use a single Microsoft Word interface tools in the Microsoft Word office software. Therefore, Microsoft independently developed VBA, VSTO, ActiveX and other technologies in user needs. Some of these technologies need to be customized and developed in the environment of office software.

Word VBA Technology

Word VBA technology automates commonly used processes or processes and creates custom solutions for existing Office productivity applications. Four software in Office all have their own programming languages called Word Basic, Excel Basic, Access Basic, Power Point Basic (the development language in Outlook is Visual Basic Scripting). Usually referred to as VBA (VB for Application) [1]. Article for Tibetan typesetting problems we can base on Word VBA technology programming and the standardization of Tibetan typesetting.

VSTO Technology

VSTO (Visual Studio Tools for Office) is an alternative to VBA that makes it easy to develop
Office applications and to develop Office applications with VSTO using many of the features in the Visual Studio development environment and the memory management, garbage collection, and other features provided by the CLR. VSTO is a set of Visual Studio toolkits for creating custom Office applications. VSTO enables users to extend Office applications such as Word, Excel, InfoPath, and Outlook with Visual Basic or Visual C#. VSTO also offers enhanced Office objects that users can program with. For example, users can also find VSTO version of Excel workbook, worksheet and range (range).

ActiveX Technology

ActiveX is a set of Microsoft's use of COM (Component Object Model, Component Object Model), making the software components in Network environment to interact with the technology set. It has nothing to do with the specific programming language. As a technology for Internet application development, ActiveX is widely used in various aspects of Web servers and clients. At the same time, ActiveX technology is also used to easily create common desktop applications. In addition, ActiveX is also a collection of open technologies that cover all popular Internet standards, languages and platforms.

Word VBA Technology Superiority

VBA technology is the customization and development of functional modules in the environment of office software as required, simplifying complicated work and facilitating repeated work. Users can customize the interface environment, create a custom function, and realize the office software built-in function which is unavailable function. The advantages of using Word VBA program to process Word documents are: ① A simple VBA program can achieve a large number of repetitive operations. ② VBA integrated in the Office software, we can directly use the functions in the Office software of VBA integrated. ③ VBA editor provides a large number of controls and a complete language system, and users can write their own needs VBA program to create a fully functional system. ④ Compared with other programming languages, VBA is a simple programming tool. Over 60% of the code can be generated by recording macros and provide visual design tools [5]. VBA, also known as a macro program, has similarities with macro. Macros are a bunch of instructions that let Office components automate the combination of user-specified actions to automate complex operations [3]. In summary, users can manipulate the content of the document as they need and can automate, generalize, normalize, and interact with each other by using the VBA program. On the other hand, the VBA program has the advantage on running in VB with a slight change.

Paragraph Layout and Technical Route

According to the Tibetan writing standard, the cursor at the end of the paragraph will be moved to the beginning of the paragraph, then the text of the current paragraph will be selected to determine whether the current insertion point falls into the form or not. If the form is entered, the format adjustment will not be performed. Manually adjust the standard format is more convenient, while paragraph layout only plain text format, the following steps.

First step: Delete the first syllable symbol. Determine whether the current line is the end of the paragraph line, then move to the first line of the line and insert point to the end of the line and select the entire line and find the text of the hit enter symbol. If you find the program can not be forward, the current paragraph is over.

Second step: Then move to the top of the line, insert the point to the end of the line, and select the entire line, hit the carriage return symbol, the current paragraph is over.

Third step: Then determine whether the current line is right-justified, the alignment conditions include a space symbol. If there is space on the right line of this line alignment, we need to find whether there is a space in this line after the right-justified.

Fourth step: After processing the clause symbol without a space, add a space after the clause
symbol, and then achieve right-justified. Specifically, we can find the Tibetan clause symbol in this line, namely the terminal symbol “།”, called “ཤད།” by Tibetan, and add a space after the clause symbol, and then we achieve right-align this line of text.

Fifth step: Handling situation which is not right-justified continuously. If there is no terminator in a line, the line can be inserted a point at the end of the line. Specifically, we move the cursor by a character to right and take the end of the word, inserting 10 points syllabic symbols to achieve the end of the line. The number of left to add is more than the length of the added characters, and move down a line continuously. The operation steps are specifically illustrated in the flowchart of Fig. 1.

![Figure 1. VBA Technology Flow Chart.](image)

**Technical Realization**

After program is programmed and the program is executed, if the cursor is not at the beginning of the paragraph when the paragraph typesetting is canceled, it will move to the beginning of the paragraph and get the current character. If the sub-syllable symbol is found, the spaces and punctuations following it are deleted, the redundant characters after the sub-syllable symbol are deleted, and the case after the clause symbol is processed. If a carriage return symbol is found, it will end the run and move one character to the right. If a clause symbol is found, the case after the clause symbol is processed. If you find the carriage return symbol, then end of the run, and move a character to the right and access to the current character to achieve programming section to withdraw paragraph layout. The implementation of the core programming code is as follows:

Step-by-step procedures are used for the progressive implementation of the program, code show as below:

1. **The first step:** Delete the first syllable symbol
   
   char_set = ChrW$(3851)   Syllable symbol
   
   DeleteInBegin (char_set) Delete the first syllable symbol
   
   2. **The second step:** Determine whether the current line is the end of the paragraph line
   
   HomeKey unit: = wdLine Move to the beginning of the line
   
   EndKey unit: = wdLine, Extend: = wdExtend Insert the point to the end of the line and select the entire line
   
   Find. Text = ChrW$(13) Enter symbol vbCr=ChrW$(13)
   
   If (.Find.Execute(Forward:=False) = True) Then
   
   bLoopCtrl = False
   
   MsgBox " The current paragraph is over "
   
   Exit Do
   
   End If
   
   3. **The third step:** Judge whether the current line is right-justified and right after the end of alignment with including a space symbol
   
   Find.Text = " " Find whether there is a space in this line, if there is space, the text is right-aligned
If (Find.Execute(Forward:=False) = True) Then
  ParagraphFormat.Alignment = wdAlignParagraphJustify
  Align this text to the right
  GoTo MyNext
End If

The fourth step: There are no spaces after the clause symbol. Add a space after the treatment of the clause symbol, and then achieve the right alignment
  Find.Text = ChrW$(3853)    '0x0F0D Find the Tibetan clause symbol in this line
  If (.Find.Execute(Forward:=False) = True) Then
    .MoveRight wdCharacter, 1  Move the cursor one character to the right
    .InsertAfter " "      Add a space after the clause symbol
    .ParagraphFormat.Alignment = wdAlignParagraphJustify  Align this text to the right
    GoTo MyNext
  End If

The fifth step: Processing continuous lines (Can not be right-aligned)
  .EndKey unit:=wdLine  Insert point to end of line
  MoveLeft wdCharacter, 1  Move the cursor one character to the right
  char_end = .Range.Characters(1)  Take the tail word
  If (char_end = ChrW$(3851)) Then        '
    InsertAfter str_end  Insert 10 sub-syllables symbols and meet the end of line requirements
    MoveLeft wdCharacter, Len(str_end) + 1 The number of left movement of characters is greater than the length of the added
  End If
  MyNext:
    MoveDown unit:=wdLine  Move down one line
  Loop
End With

Conclusion

With Word VBA programming the standardization of Tibetan layout can also partially automate the preparation of the standard. The article based on the standard writing requirements not only achieve the standard preparation of "one-click", but also can quickly achieve the layout of Tibetan documents neat. The "one-click" can quickly realize the typesetting of Tibetan documents restore and withdraw this program immediately at the same time. It realizes the incompatibility on the right side of many informational manuscripts in Tibetan newspapers and periodicals, books, essays, and the Buddhist scriptures after they were typed in Word2010 documents and enables the Tibetan documents to basically achieve the unification in typesetting. However, there are many loopholes in the process of compiling programs based on VBA technology. For example, if the compiled Tibetan program does not appear in the input Tibetan characters, the compiled program can not solve the problem align right. Besides, it has brought some limitations to standardized Tibetan version in the inconsistencies between the two terminal symbols at the end of a sentence and the existence of a space between two terminal symbols. We hope that technology compiler related procedures based on Word VBA can be resolved in the future research.

Acknowledgement

Fund Project: Innovation Group Project in Gansu Province: Intelligent Information Processing of Tibetan Language (Z16088).
References

