

## Unicode Text Editor for Ancient Egyptian Hieroglyphs Writing System

Dhamyaa A. AL-Nasrawi , Hashem A. Hashem, Mohammed A. Odhaib

Department of Computer, Collage of Science, Kerbala University, Karbala, Iraq

[dhmyaa@gmail.com](mailto:dhmyaa@gmail.com)

---

### Abstract

A writing system as a set of visible used to represent units of language in a systematic way. Egyptian hieroglyphs were a formal writing system used by the ancient Egyptians that combined logographic and alphabetic elements. In serious music textbook editor programs there is a trouble in writing documents which include Egyptian hieroglyph symbols which take more than two bytes, because there is no way to embed these symbolization in a particular document. In this paper, a special text editor designed for ancient Egyptian hieroglyph writing system which power comprehension the Egyptian hieroglyph symbols using Unicode standard, and the basic operations of a classical text editor such as operations of file manipulation (load and save), select font size, type, and color, the operations of copy, paste, cut, select all, etc. as well as print the final document. Besides this, hieroglyphic numbers were represented in this editor. More facilities, flexibility, simplicity and benefits can be gotten using these editors especially for historic research worker which they interested authors in the ancient civilization. This editor designed in Visual Basic.Net.

**Keywords:** *Egyptian hieroglyphs symbols; Unicode standard; writing system; text editor.*

---

### 1. Introduction

Writing system alludes to a way that groups of scripts may be characterized by how they speak to the sounds or expressions of human language. Case in point, the written work arrangement of the Latin script is alphabetic. In different places in the standard, “writing system” alludes to the way a specific language is composed [1].

Historically, writing systems have created after a spoken language has been made, despite the fact that the unique images utilized (commonly an ideogram) may have gone before the spoken word. An uncommon set of images known as punctuation is utilized to support in structure and association of numerous writing systems and might be utilized to help catch subtleties and varieties in the message's implying that are conveyed verbally by signs in timing, tone, stress, emphasis or sound [2].

Hieroglyphic writing is a system that utilizes characters as pictures. These distinct signs, called hieroglyphs, may be read either as pictures, as symbols for pictures, or as symbols for sounds. The aim of this paper is to design special text editor for ancient Egyptian hieroglyphs writing system which ability inclusion the Egyptian hieroglyphs symbols using Unicode standard, as well as the basic operations of a classical text editor. Rest of this paper is organized as follows. Related works explained in section 2, Section 3 introduces details of Unicode standard, Section 4 presents a definition of text editors, Proposed Text Editor for

Ancient Egyptian Hieroglyphs Writing System is explained in section 5, and Section 6 demonstrated the GUI of proposed text editor. Finally conclusions are provided in Section 7.

## 2. Related Works

There are various looks into had officially investigated in composing arrangement of antiquated dialects . Hilal M. *et al.* proposed new half and half written work supervisor for Cuneiform dialect with three classifications of symbols (assurian, Sumerian, Akkadian) . Any image might be embedded as picture in suitable size as stated by content font size [3]. The Cuneiform essayists precisely cut records of business arrangements and other commonplace matters into tablets of wet dirt. These tablets were then heated hard, in the same way as ceramics, to protect the written work. Exceptionally imperative records, in the same way as Hammurabi's Code, were frequently cut specifically into hard stone. Subodh K. *et al.* developed and distributed extensive advanced libraries of these aged tablets of wet dirt. They likewise introduce visualization prerequisites and calculations for processing them [4]. Tatsuo M., developed the familiar writing system for pictographs by utilizing the relationship to the Japanese writing technique [5].

## 3. Unicode Standard

### 3.1. Definition and Motivations of Unicode

Historically, “Unicode” refers to the Unicode Standard which defines character encodings for the writing systems of the world. It is an encoding scheme developed by the Unicode Consortium incorporated under the name Unicode, Inc. in 1991 and the ISO. It is a non-profit organization founded to develop and promote use of the Unicode standard, which specifies the representation of text in software products and modern standards[6]. Computers internally work on numbers. This means that characters need to be coded as numbers. A typical arrangement is to use numbers from 0 to 255, because that range fits into a basic unit of data storage and transfer, called a (8-bit) byte or octet[6].

It would be impractical, and often impossible, to define 256-character codes for all the possible language combinations. The solution is called Unicode[7]. Unicode's goal is to solve this issue by using more than one byte for each character[8]. The organization of Code points may range from 0 to 0x10FFFF (= 1,114,111) which divide this range into 17 planes, numbered from 0 to 16. Of these 17 planes, only 6 are currently “populated”, see Fig. 1[9]:

Plane 0 or the BMP (Basic Multilingual Plane), corresponds to the first 16 bits of Unicode. It covers most modern writing systems.

Plane 1 or the SMP (Supplementary Multilingual Plane), covers certain historic writing systems as well as various systems of notation, such as Western and Byzantine musical notation, mathematical symbols, etc.

Plane 2 or the SIP (Supplementary Ideographic Plane) is the catchall for the new ideographs that are added every year.

Planes 3-13 unassigned.

Plane 14 or the SSP (Supplementary Special-Purpose Plane) is in some senses a quarantine area. In it are placed all the questionable characters that are meant to be isolated as much as

possible from the “sound” characters in the hope that users will not notice them. Among those are the “language tag” characters, a Unicode device for indicating the current language that has come under heavy criticism by those, the author among them, who believe that markup is the province of higher-level languages such as XML.

*Planes 15 and 16* are Unicode’s gift to the industry: they are private use areas, and everyone is free to use his/her code points in applications, with any desired meaning [9].

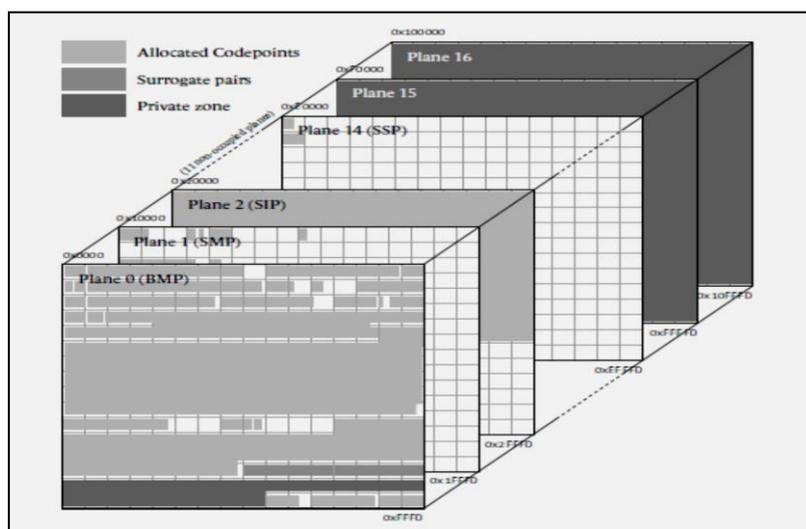


Figure 1: Planes of Unicode

### 3.2.Egyptian Hieroglyphs

Hieroglyphs (example of logosyllabaries writing system) were used to write Egyptian for more than 3,000 years, retaining characteristic features such as use of color and detail in the more elaborated expositions. Throughout the Old Kingdom, the Middle Kingdom, and the New Kingdom, between 700 and 1,000 hieroglyphs were in regular use. During the Greco-Roman period, the number of variants, as distinguished by some modern scholars, grew to somewhere between 6,000 and 8,000.

Hieroglyphs were carved in stone, painted on frescos, and could also be written with a reed stylus, though this cursive writing eventually became standardized in what is called hieratic writing. Egyptian hieroglyphs represent people and animals, parts of the bodies of people and animals, clothing, tools, vessels, and so on[1].The Unicode Block 'Egyptian Hieroglyphs' shown in table 1:

Table 1: Unicode block Egyptian Hieroglyphs

Unicode Block	
Description	Egyptian Hieroglyphs
From	U+13000
To	U+1342F
# of Characters	1071

For showing an Unicode character, it must be available in the picked font [10]. The accessibility of a particular character depends on its presence in the specified font; each font has its number of characters, or even none whatsoever. Most characters won't be accessible. On the off chance that a sought character is not display in the accessible fonts, a suitable font ought to be introduced on the system. The character might be shown on the system now, yet in the event that it is a pretty much extraordinary one, it won't be noticeable on numerous different system. An empty box, a question mark or another replacement will be shown: ❖. The font type that support Hieroglyphic symbols is *Aegyptus* .

#### 4. Definition of Text Editor

Text editor is a straightforward and an intelligent project for making and adjusting text that permits clients to make, change, or alter plain text file. The text is regularly a report, a system or maybe information for a project [11]. It could be utilized for making machine projects, altering the source code of programming languages, altering hypertext markup language (HTML), and making website pages or web outline formats. This product is most generally utilized today for programming purposes, instead of creating documents, as was in the past [12]. Just plain text may be entered into a text editor, rather than the organized text of a rich text editor or word processor. These programs permit uncommon capacities, for example, strong, italic, and different text sizes and fonts. The editor shows just the characters visible in the file, while word processors include uncommon organizing characters which are not visible in the document. One that just acknowledges plain text does not permit the client to configuration that text or include pictures or tables. The most widely recognized characteristics accessible in these programs incorporate the capacity to cut and paste text, copy text, make a bulleted rundown, wrap the text to the following line, and undo or redo the last command. Some offer more praiseworthy characteristics, for example, the capability to set macros, or command keys, for source code editing. A text editor additionally emphasizes the capacity to peruse and alter expansive file, and open them rapidly. This is essential for some expert computer applications, as they hold files that could be excessively vast for the machine's primary memory to handle. Source code documents are illustrations of extensive file that executives or expert machine clients may need to view or alter [12].

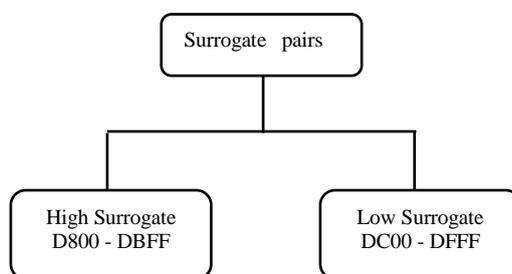
#### 5. Proposed Text Editor for Ancient Egyptian Hieroglyphs Writing System

Unicode text editor is machine programming which could be utilized to make, alter or perspective text in an assortment of alphabets. It saves data in Unicode, an advancing global standard for representation of human languages. Unicode text editor is especially suitable with non-Latin alphabets, including those that are perused from right to left and ancient scripts. Unicode editors are utilized around the world to make documents, web page substance and text for software applications in numerous languages [12]. In this paper, new text editor is proposed based on Unicode standard for ancient Egyptian hieroglyphics writing system. The problem in handling these type of writing system is how can be display symbols of these ancient language, and how to format them (by change size, color, ...etc.). The proposed text editor which facilitate the work of researchers in the ancient civilization and history field. It has more advanced features:

1. Operation of file manipulation: (create new file, open existing file, save opened file, print file).
2. Format of text: (change font size, type, and color, highlight of text).
3. Selection and manipulate of text: (copy, paste, or cut of selected text).

4. Operations of undo and redo.
5. Search operation: (find and replace).
6. Display Hieroglyphic symbols and ability to format them by change the size, color, ...
7. Represent and display Hieroglyphic numbers.

In previous works about ancient writing system, an image for each ancient symbol in a predefined size is prepared, or the ancient symbol was drawing in final historical text but this is not practical because it exhibition to errors. In this paper the ancient symbols (ancient Egyptian hieroglyphics) was formatted as a text (not as an image) using Unicode Standard. The problem here is how to display these symbols which take more than two bytes? How to overcome the 16 bit limitation? Solution is explained using surrogate pairs, this means that some 16-bit values have been reserved for use as a high (leading) or low (trailing) value in a pair of code units. Together these values denote a Unicode character outside BMP. More exactly, a high surrogate is a code unit in the range D800 to DBFF, and a low surrogate is in the range DC00 to DFFF, see figure 2.



**Figure 2: Surrogate pairs**

The steps of surrogate code unit pair algorithm are:

- Step1:** Given a Unicode code point outside BMP (with value greater than FFFF) represent it as 21-bit integer, with leading zeros as necessary.
- Step2:** Divide this sequence of 21 bits to parts with 5, 6, and 10 bits; denote the parts with  $u_1$ ,  $u_2$ , and  $u_3$ , respectively.
- Step3:** Subtract 1 from  $u_1$ , and consider the result as a 4-bit sequence. Note that this loses no information, since the original  $u_1$  is at most 10000 (because the Unicode range ends at 10FFFF hexadecimal, 10000111111111111111111111111111 binary).
- Step4:** Construct the high surrogate code unit as  $110110u_1u_2$  (by simple catenation of bit sequences).
- Step5:** Construct the low surrogate code unit as  $110111u_3$ .

Now, the Hieroglyphic symbols can be represented as Unicode code point inside BMP (with value less than FFFF) symbols, so they can be displayed.

## 6. GUI of Proposed Text Editor

It refers to the graphical interface of proposed text editor that allows users to use the features, benefits, and facilities of proposed text editor such as open and save text file as .txt, print, search about any word or replace, changing the format of text, see tools bar in figure 3.



Figure 3: GUI of Proposed Editor

The advanced feature of proposed editor is display and format the ancient Egyptian symbols and classify based on Unicode 6.2 , see figure 4 which explains two groups of Hieroglyphic symbols as example .

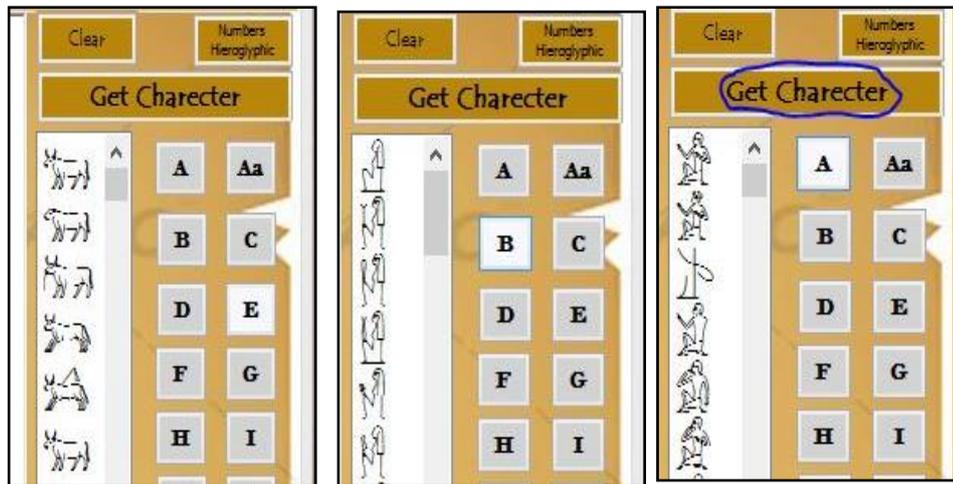


Figure 4: Two Groups of Hieroglyphic symbols

The formatting of Hieroglyphic symbols and changing their colors as example shown in figure 5.



Figure 5: Formatting Hieroglyphic symbols

New features are added in the proposed text editor by convert decimal numbers to the Hieroglyphic number see figure 6.

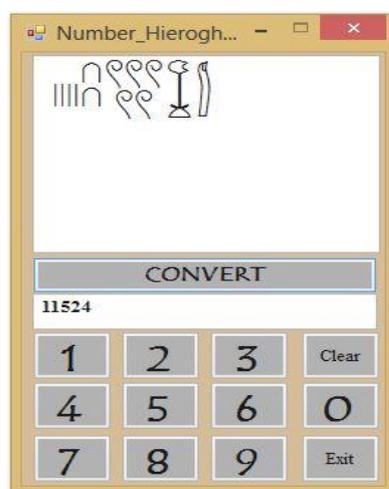


Figure 6: Convert to Hieroglyphic Numbers

## 7. Conclusions and Future Works

In proposed text editor, more facilities are produced to help user to create their documents with ancient writing system, The Unicode standard is very important way and useful tool to implement and deal with multilingual. Especially for the purpose of this paper, it was format and display ancient Egyptian Hieroglyphics writing system. Features of proposed editor and the simplicity, practical, and beautiful GUI give more facilitate to user which can get hieroglyphics symbols them self rather than image of these symbols. In the future, we can add new operations such as justify, insert pictures ...etc. and development this editor to became word processing which dealing with rich text. Another development by design Hieroglyphic-Arabic dictionary for this editor to add translation and synonym features based on [13].

## References

- [1]. The Unicode Consortium, The Unicode Standard, <http://www.unicode.org>.
- [2]. [www.en.wikipedia.org/wiki/](http://www.en.wikipedia.org/wiki/)
- [3]. Hilal M., Abdulmunim S., Haithem A. , 2006. *Hybrid writing editor for Cuneiform language*, Journal of Al-Rafidain University College For Sciences, No. 19, 1-12.
- [4]. Subodh K., et al., 2003. *Digital Preservation of Ancient Cuneiform Tablets Using 3D-Scanning*. In: Fourth International Conference on 3-D Digital Imaging and Modeling, 6-10 Oct, IEEE p. 326 – 333.
- [5]. Tatsuo M., 2010. *A Writing System for the Ancient Egyptian Hieroglyphs*. In: 7th International Conference on Informatics and Systems; 28-30 March, Cairo, Egypt: IEEE. p.1-7.
- [6]. Mohammad S. 2007. *A New Persian/Arabic Text Steganography Using “La” Word*. In: 3rd International Joint Conference on Computer, Information and Systems Sciences and Engineering; 3-12 December; Bridgeport, CT, USA: Springer Netherland. p. 339-342.
- [7]. Yannis H. 2007. *Fonts & Encodings*. Sebastopol, CA, USA: O’Reilly Media, Inc. 1040 p.
- [8]. Jukka K. 2006. *Unicode Explained*. USA: O’Reilly Media, Inc. 678 p.

- [9]. Hassan S, Mohammad S. 2006. *A New Approach to Persian/Arabic Text Steganography*. In: 5<sup>th</sup> IEEE/ACIS International Conference on Computer and Information Science; 10-12 July; Honolulu, HI, USA: IEEE. p. 310-315.
- [10]. [http://www.fileformat.info/tip/microsoft/enter\\_unicode.htm](http://www.fileformat.info/tip/microsoft/enter_unicode.htm), "How to enter Unicode characters in Microsoft Windows".
- [11]. Brian W. Kernighan, *A Tutorial Introduction to the UNIX Text Editor*, Bell Laboratories Murray Hill, New Jersey.
- [12]. <http://www.wisegeek.org/what-is-a-text-editor.htm>
- [13]. "مقار،سامح، " المعجم الوجيز(هيروغلوفاي-عربي) الخط الهيروغلوفاي في الدولة الوسطى " ،الطبعة الاولى، القاهرة ، الهيئة المصرية العامة للكتاب، 2007