Analysis of the graphetic model and improvements to the current model

Prepared by:

Badral Sanlig badral@bolorsoft.com

Jamiyansuren Togoobat onoltmn@yahoo.com

Bolorsoft LLC

http://www.bolorsoft.com

Ulaanbaatar, 2018

Abstract

We have researched both graphetic and phonetic encoding models. Every model has drawbacks and benefits. This document consists of three major parts. In the first part, the archeology of Mongolian language is introduced and in the second part, the analysis of the graphetic model is presented. The proposal of the improved phonetic model comes in the third part.

We examined that the language and script information is very important for the decision-making, thus we introduced the basic information of the Mongolian script.

We have defined the current graphetic model as semi-graphetic model, because it was extensively mixed with phonetic elements. Thus, we also analyzed pure graphetic model.

We have concentrated to improve current phonetic model with minor and major updates.

1. Introduction to the Mongolian script

The Mongolian script has been used as an official writing system since the thirteenth century. The oldest Mongolian script monument, known today, is written in 1224 or 1225, and furthermore, the first treatise of Mongolian grammar dates back to the 13th century. For example, The Mongolian grammar *Jirüken-ü tolta* (The Heart Essence) by Sakya Pandita Kunga Gyaltsan (Sa skya Pandita Kun dga' rgyal mtshan(1182-1251)), *Jirüken-ü tolta* by Chogyi Odser (Chos-gyi 'Od-ser (Choiji-Odser, fl. 1307-1321)). Unfortunately, these precious works have not been handed down to us today. However, several later day "commentaries" on these works, dated back to the eightteenth century or later, are available. For an instance, one can duly mention the *Jirüken-ü tolta-yin tayilburi üsüg-ün endegürel-i arilɣaɣci Oɣtarɣui-yin mani* (The Space Jewel for Eliminating of Letter Ambiguity: Commentary on the Heart Essence). This is the earliest and most popular commentary, which was written by Danjindagba (fl. 1723-1736), a famous reincarnated Lama from the Üjümchin Mongols. From these works one can see serve as how Mongolian script is an embodiment of ancient Mongolian orthography.

Therefore, I contend to introduce some of the Mongolian orthography terms on the basis of the Jiruken-ü tolta commentary by Danjindagba. A Reason is I regard this commentary as a most fundamental source for the sake of encoding the Mongolian script.

Based on the work of Danjindagba, I will briefly demonstrate five unique characteristics of vowels, consonants, syllable and syllable closing, which are essential in the Mongolian grammar.

Historically, the alphabetic set had been improved several times. The present alphabetic sources of the Mongolian script can be divided into 7 vowel letters and 28 consonant letters. The seven vowels are: ᠠ a, ᠡ e, ᠢ i, ᠣ o, ᠤ u, ᠥ ö, ᠦ ü. Basic consonant letters are: ᠨ na, ᠪ ba, ᠬ qa (ᠬ ke), ᠭ ɣa (ᠭ ge), ᠵ ja, ᠶ ya, ᠲ ta, ᠳ da, ᠮ ma, ᠴ ča, ᠷ ra, ᠰ sa, ᠱ ša, ᠯ la, ᠸ wa , ᠫ pa, ᠩ ang and ᠯᠬ lha. They are used to write native Mongolian words. While there were other consonant letters to write foreign words. Those are: ᠹ fa, ᠽ za, ᠼ ca, ᠆ ža, ᠻ ka, ᠬ ha, ᠽ Zhi, ᠼ Chi. There are one and the same letters for denoting o and u; and ö and ü as well. Though, they are considered as separate letters according to *the Jirüken-ü tolta.* For example, it is quite clear that Danjindagva regarded o and u, and ö and ü as separate letters. Let me illustrate three examples where he regarded them as separate letters as referring to his work. Those are as follows:

1) " ... *A generates* o *and* u. *While* e *generates* ö *and* ü. *Na generates* no *and* nu. *Ne generates with* nö *and* nü. *Ba generates* bo *and* bu. *Be generates with* bö *and* bü" (Danzandagba p.6r). Here, he attempts to explain that a "contour", which is called *gedesü* or the "belly" in Mongolian script, indicates two different letters o and u.

   Unfortunately, he had to write O, the "contour" (*gedesü* or belly) twice to mean

this. He did the same to explain the use of *ö* and *ü*. The difference between *o* and *u* and *ö* and *ü* is the former are consistent with their back vowels while the later are consistent with their front vowels only. Nevertheless, it is again the same O "contour" (*gedesü* or belly) used to indicate *ö* and *ü,* from the difference between front and back vowels one can identify them as different letters.

2) "... the vowels are *a, e, i, o, u, ö and ü*. These are named as *aq-a üsüg* (initial letters) or *eke üsüg* (mother letters) due to their first positions in writing and spelling with consonants" (Danjindagba, p.7a). Here, Danjindagba uses seven letters to indicate seven vowels, not five letters. To mean this he had to use O the circle four times.

3) "...The letter *š* occurs with vowels such as *ša, še, ši, šo, šu, šö* and *šü*" (Danjindagba p.6r). Here, he shows seven different uses of the consonant *š* in the case of seven different vowels. If he did not consider *o* and *u*, and *ö* and *ü* as independent letters, he would have described only five syllables (i.e. *ša, še, ši, šu, šü*). However, he demonstrated seven syllables to indicate that although they share the same grapheme they are independent letters.

Jiruken-u tolta underlines another important feature of Mongolian script, which is the *vowel harmony*. According to Danjindagba, there are three types of vowels in Mongolian language.

1) *er-e* or *čingγ-a egesig* (lit. masculine or strong vowel) which means back vowels - *a, o, u*.
2) *em-e* or *köndei egesig* (lit. feminine or weak vowel) which refers to front vowels - *e, ö, ü*.
3) *saγarmaγ egesig* (lit. neutral) – *i*.

The vowel harmony simply means that a word can only contain either back vowels (a, o, u) or front vowels (e, ö, ü), but not both at the same time, with the exception few of words, the majority of which are foreign. The vowel *i* is considered neutral, and therefore, it occurs in both front and back voweled words, but when *i* occurs in all syllables in the words, then the word is considered to be front voweled. Vowel harmony also affects two other sets of letters, γ/q and g/k, the former occurs only in the back-voweled words, while the latter only in the front-voweled words.

Consonants are classified into two categories. One is the syllable closing consonants that can be followed by another consonant in the middle of a word and may occur in the end of a word (debiskerlekü geyigülügči). Second is the consonants that always followed by vowels (ülü debiskerlekü geyigülügči). According to Danjindagba, there are eleven syllable-final consonants.

For example, Danjindagva writes that "... There are eleven masculine syllable-closing consonants: *an, ab, aγ, am, al, ar, as, ad, ay, aw* and *ang*. ... The feminine syllable-closing consonants are *en, eb, eg, em, el, er, es, ed, ey* and *ew*" (p. 7r). Among these syllable closing, *w* and *y* transformed to vowels *u* and *i* in Modern Mongolian. A well-known example of syllable-closing *y* is *nayma* ᠨᠠᠶᠮᠠ (eight). This unique writing form, which is written with only one long tooth or oblique line (ᠢ), and it differs from the other standard diphthong writings (ᠠᠢ). Some examples of syllable closing *w* are available such as *keüked* ᠬᠡᠤᠬᠡᠳ (child), *taulai* ᠲᠠᠤᠯᠠᠢ (rabbit), *teüke* ᠲᠡᠤᠬᠡ (history) etc.

The syllable-final *ng* is a special one. It does not occur in the beginning of the word (i.e. not followed by vowels).

Danjindagva explains that "... The letter *ng* does not occur in the first position of a word. It only occurs to describe the sound of the newborn baby crying. Therefore, it is considered as a syllable-final consonant. For example, babies cry as *ing ng* (ᠢᠩ ᠩ)."

This feature does not appear in Mongolian CV Sequences_Weizhe171209.pdf where the *ng* consonant is followed by vowels. At a glance, one can see that the author of this work wishes to not use ligatures in his model. However, this plan fails because of the feature of *ng* consonant.

The syllable structures are well described in the *Jirüken-ü tolta*. For example,

Vowel syllable / V: ᠠ ᠡ ᠢ ᠣ ᠣ ᠥ ᠥ (a, e, i, o, u, ö, ü).

Single syllables that consist of vowel and syllable closing consonant / VC (dang debisker üy-e): ᠠᠪ ᠠᠮ ᠠᠭ ᠠᠩ ᠠᠯ ᠠᠰ ᠠᠳ ᠠᠨ ᠠᠩ ᠠᠷ

Syllables with diphthongs and syllable-final consonants (dabqur debisker) / VYC: ᠠᠶᠢᠷ ᠠᠶᠢᠨ ᠠᠶᠢᠯ

Consonant and vowel / CV: ᠪᠠ ᠨᠠ ᠷᠠ

Consonant, vowel and syllable-final consonant / CVC: ᠪᠠᠷ ᠪᠠᠨ ᠰᠠᠩ ᠳᠠᠯ

Consonant, diphthong and syllable-final consonant / CVYC: ᠣᠩᡤᡘ ᠊᠊ᡘᠣᠣᠣ

Loan words such as *tngri* ᠊ᠣᠩᡰ (sky) and *gšan* ᠊ᠩᡰ (moment) are not included in this structure.

НЭГДҮГЭЭР ХИЧЭЭЛ

Давтлага:

Дэвсгэрлэх, үл дэвсгэрлэх үсэг

Монгол бичиг 32 үсэгтэй. Гийгүүлэгч 25-аас харь хэлний авиа тэмдэглэх ᴖ ᴫ ᴙ ᴣ ᴣ таваг хасвал үлдсэн 20 гийгүүлэгч нь монгол хэлний авиа тэмдэглэдэг гийгүүлэгч болно. Тэдгээр нь:

1.	(НА)	8.	(ЙА)	15.	(ША)
2.	(БА)	9.	(ТА)	16.	(ЛА)
3.	(ХА)	10.	(ДА)	17.	(ВА)
4.	(ХЭ)	11.	(МА)	18.	(ПА)
5.	(ГА)	12.	(ЧА)	19.	(НГ)
6.	(ГЭ)	13.	(РА)	20.	(ЛХА)
7.	(ЖА)	14.	(СА)		

Ингээд 7 эгшиг, 20 гийгүүлэгч-бүгд 27* үсгээр монгол хэлний авиаг тэмдэглэхэд хүрэлцээтэй юм.

Тэдгээр 20 гийгүүлэгч үсгийг 1) дэвсгэрлэх үсэг, 2) үл дэвсгэрлэх үсэг гэж хоёр ангилдаг.

> Дэвсгэрлэх буюу дэвсгэр үсэг гэдэг нь үе битүүлэн ордог гийгүүлэгч мөн.
> Өөрөөр хэлбэл: үеийн адагт ордог, үе төгсгөдөг гийгүүлэгчийг дэвсгэрлэх үсэг гэнэ. Бас өөрөөр хэлбэл: үгийн дунд, гийгүүлэгчийн өмнө ордог, үгийн адагт ордог гийгүүлэгчийг дэвсгэрлэх үсэг гэнэ. Бас өөрөөр хэлбэл: үгийн дунд, үгийн адагт ардаа эгшиггүй орж болдог гийгүүлэгчийг дэвсгэрлэх үсэг гэнэ.

Ga, Qa, Ge, He letters are taught as individual letters in the School Textbook for Mongolian Script printed in Ulaanbaatar, 1986.

2. Analysis of the graphetic model

The proposed graphetic model in N4889 is not true graphetic model. Thus, we would describe it as semi-graphetic approach.

Semi-graphetic model

We have carefully checked the advantages of the semi-graphetic model described in N4882 as follows.

1. Cleaner, unambiguous representation of text.

There exists still ambiguous representation of text.

ᠬᠠᠭᠠᠨ qagan ᠬᠠᠨᠨᠠᠨ qannan

ᠠᠭᠠᠳ /aagad/, /aannad/, /aagaon/, /aannaon/

ᠠᠭᠠᠭᠤᠷ /aaagagur/, /aqgagur/, /aaagnannor/

2. Vastly simpler font implementation, with only local contextual rules.

We accept this issue. However, we could reach to this goal with tiny changes to current model.

3. No variation sequences required for modern Mongolian (except to support old model for backwards compatibility).

We accept this issue. We could also significantly reduce FVSs in current model. Actually just single FVS character is enough.

4. More straightforward user experience; type what you see.

It is incorrect. For example, to ᠲᠣᠨᠠᠨᠨ᠎ᠠ ᠲᠣᠨᠠᠳ

It is impossible to abandon from this issue unless to avoid positional variants.

5. Much easier searching capability.

It is impossible to abandon from this issue unless to avoid visual ambiguity. This visual ambiguity is not like as phonetic model but it is emerged by mixed encoding of the model. For example, see the point 1.

6. Would be supportable in internationalized domain names.

There are still unacceptable security issues for domain names.

7. Has significantly less security issues than the status quo.

There exist still security issues.

This approach has following drawbacks:

1. Mixed (graphemes and letters) encoding causes more difficult problems to mentality. **A mongolian user cannot write neither by correct spelling nor by thinking the pure letter elements.** For example, "öndör" is correct spelling of

the word ᠥᠨᠳᠦᠷ /meaning high/ will be spelled as "aueadur". Almost all words will be spelled incorrect.

2. Retains other minuses of the graphetic approach.

Issues on migration

The migration issues of the graphetic model are described N4890. The migration of encoding was never being so simple. The coexistence cannot be disappearing. There exist still huge problems with Cyrillic script in Mongolia. Before Unicode age, we used win1251 for Mongolian Cyrillic encoding. There were just two letters with two variants for each more than Russian. From 2000 until now, there exist still coexistence problems. There are bunch of incorrect fonts, which contain Ө,Y letters both in ANSI block and Cyrillic block. (0400-04FF)

Thus, the semi-graphetic approach is unacceptable model.

Further, we analyzed pure graphetic model.

Pure graphetic model

The ideal graphetic approach is the model of old typewriter shown as in Figure 1.

Figure 1

This graphetic model has been used until the mid-90's.

Every Mongolian letter consists of defined strokes (not grapheme).

This is the traditional approach for Mongolian typewriting system. Figure 1.

Various elements of letter

According to Mongolian linguistic books, there are specific terms to describe the elements of a letter. These terms they represent, are often referred to as parts of letters or positional variations of the graphemes. These pedagogic terms divided into following three groups: elements of letter or constituent parts of letter, syllable closing consonants and enclitics.

There are basic eleven elements of letter (mo. ᠵᠤᠷᠠᠮ *zuram*). All these names were derived from shape and position of the elements.

- ᠠᠲᠰᠠᠭ **Atsag /** Aleph, teeth-denote
  Horizontal diamond shaped stroke of A, E, NA, GA, DA, NG, MA, LA; known as teeth or tooth.

- ᠲᠢᠲᠢᠮ **Titim /** Crown
  Styles are often divided into with or without titim(analogy to serif and sans serif). Titim style is distinguishable by initial part of the all seven vowels, some consonants and suffixes. It marked consonant weakening. This element does not contain any phonetic information. Titim was first appeared in the 19th century as denoting sound weakening consonants.

- ᠱᠢᠯᠪᠢ **Shilbi /** Long teeth
  A long bar is diagonal line of letter I, JA, YA, also known as an identification sign for feminine vowel.

- ᠡᠸᠡᠷ · ᠭᠡᠵᠢᠭ **Ever, Gezeg /** Horn
  A diacritic marks used in MA, CA, ZA, LA,

- ᠰᠡᠭᠦᠯ **Suul /**Tail
  Long or short strokes, which usually written at the end of the words.

- ᠭᠡᠳᠡᠰ **Gedes /** Contour
  A closed circular line that creates interior space, such as O, U, OE, UE, BA, PA,

- ᠨᠤᠮ **Num /** Bow
  An open circular line, such as in KE, GE, NG, BA, PA and FA.

- ᠨᠢᠷᠤᠭᠤ **Nuruu /** Backbone
  There are two different meanings for this term. First, it is a whole body of the word linked in one line. Second, a space that is inserted between two letters.

- ᠵᠠᠪᠵ **Zavj /**
  An angular open contour, such as SA and SHA.

- ᠵᠠᠷᠲᠢᠭ **Zartig /**
  A component of in shape of ear.

- ▰▰ **Tseg,** ◦◦◦◦◦ **dusal** / Point
- A Drop like diacritic mark which is used with N, G, Sh.
- ◦◦◦◦ **Shavj** / Finial
- A terminal, a term that is invented by printing professionals for its convenient typeface usage.

Some of **Zuram** could be called as positional form of grapheme. For an example: A longer tail is a final form of A, E and N. A crown is an initial form of E.

Above-mentioned elements of letters are commonly taught in school in order to give detailed information of the graphic of letters.

Lubsanbaldan, 1972, x. 209–219; Лувсанбалдан, 1975; Кара, 1972, x. 41; Мижиддорж, 1976; Шагдарсүрэн, 1975; 1984; 1987]. Нэр томъёог ерөнхийд нь зурлагын нэр, авиа дуудлагын нэр, дагавар болон сул үгийн нэр гэсэн гурван том бүлэгт хуваяж болно.

Зурлагын нэр

1. *ačuγ* (•) — ацаг
2. *sidü* (•) — шүд
3. *örgesü* (•) — өргөс
4. *niruγu* (—) — нуруу
5. *γoul* (—) — гол
6. *silbi* (•) — шилбэ
7. *em-ün temdeg* (•) — эмийн тэмдэг
8. *urtu sidü* (•) — урт шүд
9. *segül* (•) — сүүл (•• үсэгт)
10. *degegsi ebertei silbi* (•) — дээш эвэртэй шилбэ
11. *doγuysi ebertü silbi* (•) — доош эвэртэй шилбэ
12. *eteger silbi* (•) — этгэр шилбэ
13. *erteger silbi* (•) — эртгэр шилбэ
14. *yatuγar silbi* (•) — ятгар шилбэ
15. *erbegeljin silbi* (•) — эрвээлжит шилбэ
16. *gedesü* (•) — гэдэс
17. *qoduγudu* (••) — ходоод
18. *baγa qoduγudu* (••) — бага ходоод
19. *yeke qoduγudu* (•) — их ходоод
20. *bitegü* (•) — битүү
21. *γoγčuγa* (•) — гогцоо
22. *gejige* (ᚠ) — гэзэг
23. *eber* (ᚠ, ᚠ) — эвэр
24. *degegsi eber* (ᚠ) — дээш эвэр
25. *doγuysi eber* (ᚠ) — доош эвэр
26. *titem* (•) — титэм
27. *qaγarqai toluγai* (ᚠ, ᚠ) — хагархай толгой
28. *angarqai toluγai* — ангархай толгой
29. *aγsabur* (ᚠ) — агсвар
30. *numu* (ᚠ) — нум
31. *jabaji* (ᚠ, ᚠ) — завьж
32. *ereü* (ᚠ, ᚠ) — эрүү
33. *aru-yin sa* (ᚠ, ᚠ) — арын са[1]
34. *öbür-ün sa* (ᚠ, ᚠ) — өврийн са[2]

---

[1] Ихэвчлэн буриад уламжлалд хэрэглэдэг нэр.
[2] Ихэвчлэн буриад уламжлалд хэрэглэдэг нэр.

Figure 2

More than 100 characters/elements are necessary on keyboard for pure graphetic model to illustrate all mongolian letters including ali gali letters.

There exist also visual ambiguities. We concluded that for mongolian script, we never abandon visual ambiguities.

With this model, we cannot store and transmit Mongolian language information and we suspect graphetic model cannot live long due to usability and text processing problems.

Thus, the graphetic model is also unacceptable.

3. Improvements to the current model

Analysis of the phonetic model

As mentioned in first chapter Mongolian script was unmistakably phonetic model. Choiji Odser, linguistic scientist in 13th century, was written primary source of current Mongolian script model *Jirüken-ü tolta* (Aorta of the heart), which is adapted from Uigur script. Even though the original publishing of "*Jirüken-ü tolta*" is not found, besides *Jirüken-ü tolta-yin tayilburi Oɣtaruyin mani* (Mantra of the space: Commentaries on the Aorta of the heart), of Danzandagva is found. In this source, issue of the phonetic and graphetic is also noted/appeared in that time and followed the phonetic approach.

For that reason, we continued the research of issue on phonetic approach is still compromising.

Phonetic model is almost unmanageable near in the future. Why?

1. Incorrect and overloading usage of FVSs, which resulting unnatural user experience.
2. Heterogeneous implementations of fonts without clear specification.
3. Uncovered font rules and some missing characters.

Thus, it is necessary to eliminate all disadvantages of the current model that are mentioned in N4882.

Encoding issues

We have carefully checked the Mongolian block from initial standard until current standard (Unicode 10.0) as well as technical report 170 to determine whether the original model proposed in Unicode 3.0 incorrect.

We found some critical problems or mistakes like incorrect encoded Mongolian letters QA, GA, which play main role to control vowel harmony rules of Mongolian script. Currently, we are working exactly reversed way. That is we always tried to determine those letters from context by font rules. We proved it is impossible. Currently, we could

not distinguish masculine and feminine form of those letters even though we wrote significant number of rules.

We could considerably reduce the complexity and contextual rules by strictly isolating Mongolian letters QA - QE, GA - GE.

- NNBSP is not well defined and intended. Please see the proposal of L2/17-036.
- In recent version of the standard, the stylistic and periodic characters are encoded. They have to be cleaned up.
- There are some missing characters in Mongolian block.
- There are missing character sets in Mongolian Ali Gali block.
- There are missing characters in Mongolian TODO block.

Free Variation Selectors

Locating these control characters on input device is significantly confusing users and untypical experience for users. As I see we don't have to place more than one free variation selector on input device. For character variant selection we have a good exemplar. There was a good typewriting software (in DOS environment) namely "Sudarch" in 90's. This editor had a free variation selector key which used for all variants.

For instance, analogue to current Unicode FVS1 used one FVS, for FVS2 two sequential FVSs and for FVS3 three sequential FVSs. The advantages are:

a) User does not need to search where FVS keys are located on keyboard.

b) User can directly see which variant is displayed by which variation keys (1, 2, 3 etc.) by typing FVS or Backspace keys.

Narrow No-Break Space

NNBSP causes many problems. First of all, we cannot use it as suffix connector. Because almost all plattforms didn't support this character except microsoft. In most systems, NNBSP is replaced by SPACE (0020). We could prove that easily. For instance, try to type in Facebook messenger or copy and past NNBSP contained text to the messenger.

Currently, the active users use space and nirugu to illustrate the suffixes correctly.

Some missing characters
Quotationmarks

For quotationmarks, we use latin characters << >> because almost we don't use CJK fonts. The problem is <<>> seems ugly and not centered. It's required to add this punctuation. For other punctuations we use latin characters but they doesn't placed in

the middle. Should we redraw all required lating glyphs to use them? We need clear instruction for font designing.

Abbreviations of first letters

It's necessary to add a full stop sign without a space after that to write abbreviations like surname etc.

Composite word joiner

We need one more control character to write joined words like Batumungkhe (ᠪᠠᠲᠤᠮᠦᠩᠺᠡ), Ueruntuyaga (ᠦᠶᠡᠷᠦᠨᠲᠤᠶᠠᠭ᠎ᠠ), GereltOd (ᠭᠡᠷᠡᠯᠲᠣ᠋ᠳ). This method is used to avoid the confusions like GerelTod (ᠭᠡᠷᠡᠯᠲᠣ᠋ᠳ). To distinguish composed words we use traditionally a long nirugu. It is currently possible using two nirugu and FVSs but it is difficult from user experience point of view. If it's possible a nirugu like character which has an effect following syllable.

Input device

We need to show some characters to aware of invisible characters like FVSs are more than one times typed if they are placed on input devices.

Currently, the users are unable to know how many times they typed invisible characters.

The demonstration will be presented at the meeting.

Font

Major problem of current encoding is overloaded rules in fonts that they destabilize the encoding. We can easily find numerous examples with calt rules of QA, GA letters and instable FVSs like ᠪᠢᠴᠢᠭ /bicigVFS1/, ᠪᠢᠴᠢᠭ /bicig/

The best way to stabilize FVSs, we need to standardize the rules of the font and the rendering.

See appendix.

The improved phonetic model

The proposed improvement of the current phonetic model could be realized in ways.

Minor changes:

1. Fixing positional mismatches
2. Minimize Free Variation Selector at least on input devices
3. Clean up the stylistic and periodic characters and reorganize some variations into ali gali section.

4. Separately encode MONGOLIAN LETTER QA, MONGOLIAN LETTER GA as fixed variant with FVS1.
5. Standardize miscellaneous font rules (OTF)
6. Standardize rendering rules
7. No radical changes

Major changes:

Following changes has to be made in addition to minor changes.

1. Encoding MONGOLIAN LETTER QA (182C) separately by feminine and masculine forms.

2. Encoding MONGOLIAN LETTER GA (182D) separately by feminine and masculine forms.

We have to make a decision with cost-effective, efficient and long-living technologies designed to meet our requirements. Should we consider next 20 years or should we 2000 years?

Should we store and transmit physical real language information or highly abstracted visual information?

On the basis of this analysis, we concluded that the current model is most outstanding encoding model of Unicode.