Search Interface to a Mayan Glyph Database based on Visual Characteristics[*]

Grigori Sidorov[1], Obdulia Pichardo-Lagunas[1], and Liliana Chanona-Hernandez[2]

[1]Natural Language and Text Processing Laboratory,
Center for Research in Computer Science (CIC), National Polytechnic Institute (IPN),
Av. Juan Dios Batiz, s/n, Zacatenco, 07738, Mexico City,
Mexico
sidorov@cic.ipn.mx, ayilina@hotmail.com

[2]Superior School of Mechanical and Electric Engineering (ESIME),
National Polytechnic Institute (IPN),
Av. Juan Dios Batiz, s/n, Zacatenco, 07738, Mexico City,
Mexico

Abstract. During years, the community of Mayan researchers was not open to the usage of computer tools. Still, the progress of the computer science and the current state of Mayan research proves the necessity of this type of software. We present the project related to the development of Mayan script database, which is the first necessary step in development of computer representation of Mayan script data. The database contains several tables and allows for various queries. The main idea of the project is the development of the system that would allow managing Mayan script data for specialists and as well for persons without any previous knowledge of Maya. The system includes structural visual description of glyph images and search facilities based on the structural description, along with traditional database search facilities. The user can define any set of visual characteristics. The glyphs are characterized using a "naive" feature set. Fully working prototype is presented. Preliminary evaluation of the efficiency is carried out. In fact, the system can be applied to any set of images that can be assigned structural features.

Keywords: Search interface, Mayan glyph database, visual structural description.

1 Introduction

Mayan hieroglyphic writing contains more than thousand of different glyph signs. Many of them are variations of one sign (allographs), other glyphs are different signs with the same reading (homophones). Some signs are glyph variants that were used during a certain period of time or in a certain area.

Mayan writing system can be described as a logosyllabic system ([1, 3, 4, 5, 7]), based on signs that represent words (logograms) and syllables that also can be used as phonetic signs. There are approximately 200 different syllabic (i.e., purely phonetics) signs, of which about a 60 percent are homophones. Namely, there are about 80 syllables in the classic Mayan language according to its phonetics; still more than 200 glyphs signs were used in the phonetic writing. The Mayans used a system of writing capable to register complete oral manifestations of their language.

The discoveries of the last decades in Mayan epigraphy field allowed deciphering of almost all documents and known inscriptions, according to the dominant theories of deciphering. Though, the deciphering is different according to the adopted theory. Nevertheless, there is no kind of somewhat sophisticated computational tool that can be used by the Mayanists in their investigations. These investigations are normally carried out manually on the basis of facsimiles of documents and Mayan inscriptions ([6, 8, 9]).

On the other hand, the public in general should also have possibilities of using some specific computer tools if they are interested in reading Mayan inscriptions. Though they are not specialists in the field, the subject is interesting from the general cultural perspective.

Even though the specialists are aware of the existence of various dictionaries in the Internet, they are not frequently used. Even more: during years, the community of Mayan researchers was very skeptic as far as the usage of any type of computer tools. Still, the progress of computer science and the current state of Mayan research proves the necessity of this type of software, because it makes the work of a researcher more fast, reliable and productive.

We present the project related to the development of the interface to the Mayan glyph database, which is the first necessary step in development of computer representation of Mayan script data.

We believe that developing of this application would be impossible without support of investigators of Maya and without considering their needs as end users. Our purpose is development of the application in which investigators of Maya would participate in design and implementation. Also, this application should be able to resolve specific problems that can solely be raised by a specialist.

This software will help to diminish the burden of many rather complex procedures that are performed manually by now, like context comparison or glyph identification. In addition, this software offers to the investigators a tool that facilitated the process of search and classification of Mayan glyphs and can serve as a didactic tool that helps in learning of the glyph signs that compose Mayan writing system.

The paper has the following structure. First, we describe the interface and its usage (Section 2), then in Section 3 we present the preliminary evaluation of the interface based on the comparison with the purely manual search, and finally the conclusions are drawn.

2 Interface Description

We based our development on John Montgomery's dictionary [2]. It contains hieroglyphs of Mayan classic writing, organized alphabetically by phonetic writing of words, phrases or syllables, and it also includes appendices that list signs using additional classification categories.

The developed application is a relational database – and the corresponding interface – that stores general information about glyphs: glyph image, translation, transcription, Thompson numbers (these can be considered as glyphs identifiers, ID; they were assigned in arbitrary manner by E. Thompson, an outstanding Mayan investigator), phonetic reference and descriptive notes corresponding to each glyph that forms part of John Montgomery's dictionary. Total size of the dictionary is 1,241 entries. The database is ordered according to Thompson numbers, but the system offers two more options for its ordering using the fields containing transcription or translation.

Fig. 1. Interface for the database (main dictionary view).
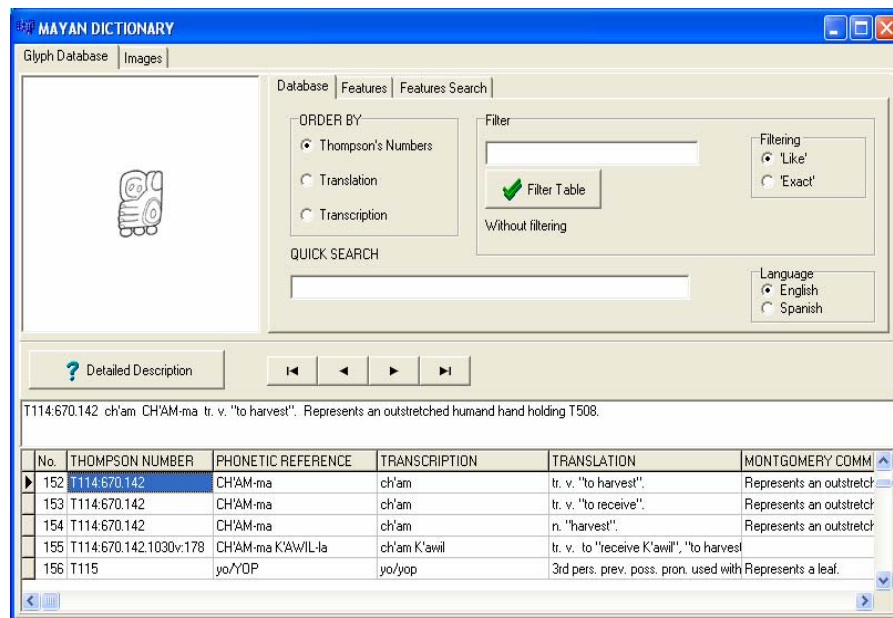
Due to normalization of the database, in the cases that a glyph has more than one meaning according to the dictionary, these are separated and marked with ID that corresponds to the sign and as well with the IDs that correspond to the number of meaning (translation); thus, different meanings are stored as separate records in the database.

The database is also capable of storing information provided by a user related to the structural visual description of the graphical image. This is based on a feature set

created by each user if he wants to describe the glyphs using his own feature set, see below. The system has the predefined feature set. The user has a possibility to choose a desired feature set: the predefined one or his own set.

The information about glyphs is visualized as records in a standard tabular way, i.e., each visual field value corresponds to a specific database field. The following fields are used: glyph image, Thompson number, phonetic reading, meaning (translation from Maya), and comments.

Different meanings are stored as separate records (Fig. 1), thus, duplicating other field values. We chose this option for enabling searches in the field "meaning". Still, we have only one glyph image visible at a time and we present additionally the concatenated information of values referring the meaning (from several records) just below the glyph. This is justified by the fact that otherwise this information is presented in separate records and, thus, it is not clearly related to the same glyph.

The complete information about glyphs can be visualized within separate dialog window, see Fig. 2.

Fig. 2. Complete information about a glyph.
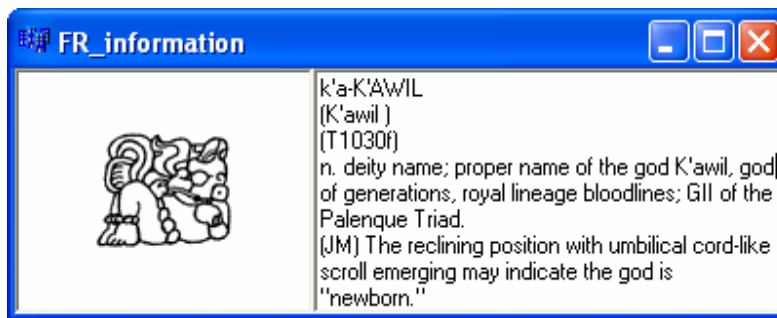
The user has an option to choose the ordering of the dictionary records by marking the corresponding field in the Button with multiple choices "Order by": *Thompson number, translation* or *transcription*.

This interface also allows for two types of searches: fast search and filtering. Both of them are performed in accordance with the selected ordering. The fast search moves the table cursor to the desired value according to the ordering in the corresponding field. It starts working with partial matching, thus, the cursor is moved to the first value that matches the current search field (though partially). The filtering restricts the records that are shown to the user according to the data present in the field "Filter" and the current ordering.

For the advanced search using structural features, the user can use predefined feature sets or create his set and describe glyphs according to this set.

The systems shows the window, where it provides the general information about a glyph. Also, a list of possible graphical features (characteristics) is presented, see Fig. 3.

The system allows for structural description of glyph images using various feature sets. This facility permits a user who is not familiar with Maya glyphs to make searches and identify the glyphs. The user can add and remove these features while

describing an individual glyph. This makes possible assigning characteristics that correspond to each one of the glyphs. Each user (professional or amateur) can use his judgments for this characterizing.

Fig. 3. Feature set assignment.
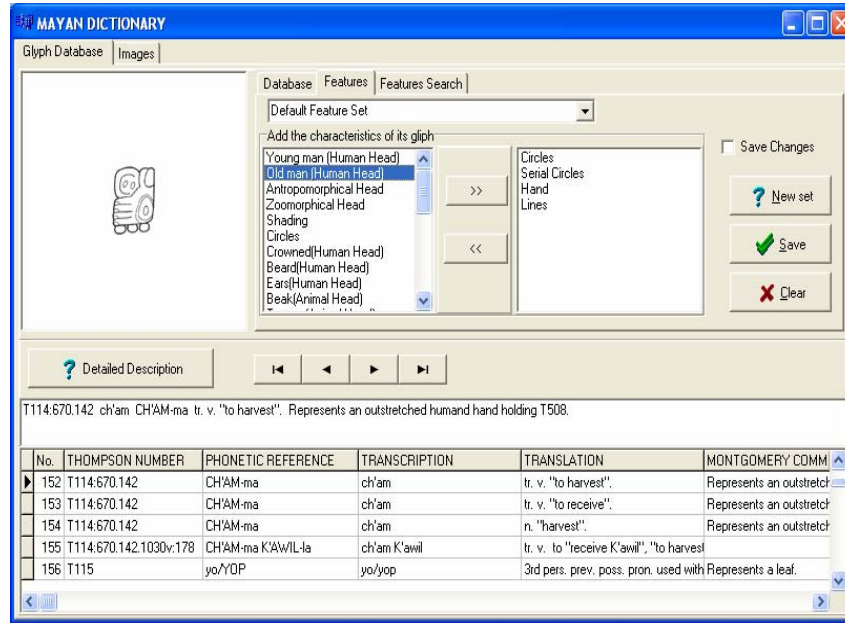
The system stores each one of assigned feature values and allows visualizing of the characteristics that have been assigned previously to the glyph. Each set of features has its own name, thus, changing the set is just picking up a different set name. This allows for different users applying different feature sets prepared by them or by other users.

When the user wants to search for a glyph, first he should choose the corresponding feature set, see We made several searches using our feature set. Clearly, the search time of the specific glyph is reduced if we filter out not relevant glyphs. As far as the number of filtered glyphs is concerned, for the most "specific" glyphs, the results contained 5-10 rows. For the most "typical" glyph (for example, if we select the feature "circles" only), the result contained about 300 glyphs, which is much less than the total 1,241 glyphs in the dictionary, but still it is too many for manual scanning. Still, it is the worst case, and the user is expected to select at least one more feature, that reduces the search space at least three-four times.

Fig. 4. In this case, the set presented is "Default feature set", i.e., the "naive" features that we assigned to all glyphs. The set of these features is presented in **Fig. 5**. The user can characterize the desired glyph using the features of the feature set using this dialogue. If there is no exact matching, the system will ask a user to choose

possible additional features that correspond to the set of features of real glyphs from the database, see **Fig. 6**. This reduces the search space even more.

3   Evaluation of Interface Efficiency

As we already mentioned, we characterized all glyphs using the "naive" feature set. Using this feature set, we made several experiments related to the functioning of the system. Obviously, the results depend basically of the way we define our feature set. This is a future work that should be performed basically by the specialists in Maya writing.

We made several searches using our feature set. Clearly, the search time of the specific glyph is reduced if we filter out not relevant glyphs. As far as the number of filtered glyphs is concerned, for the most "specific" glyphs, the results contained 5-10 rows. For the most "typical" glyph (for example, if we select the feature "circles" only), the result contained about 300 glyphs, which is much less than the total 1,241 glyphs in the dictionary, but still it is too many for manual scanning. Still, it is the worst case, and the user is expected to select at least one more feature, that reduces the search space at least three-four times.
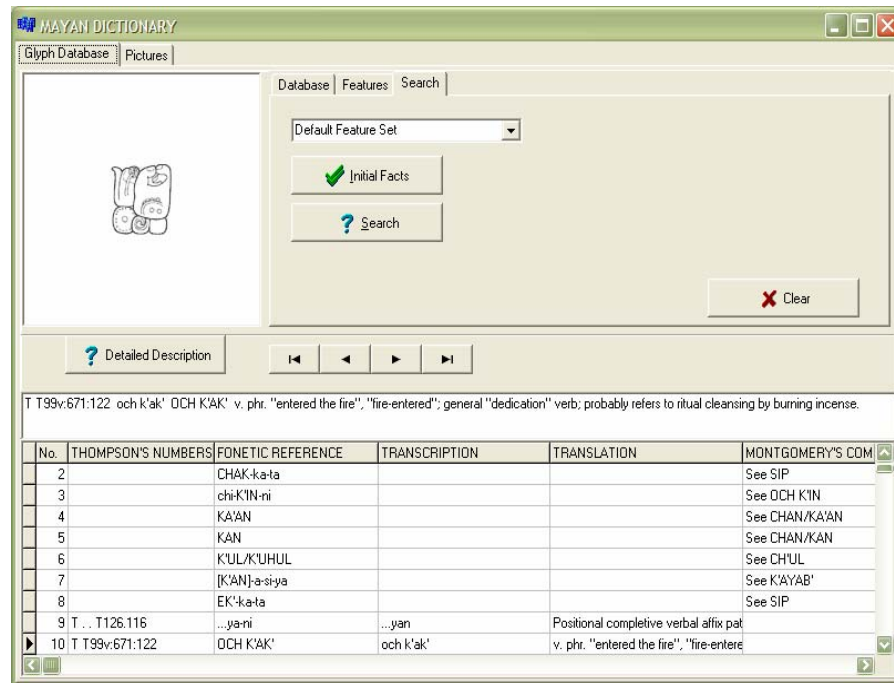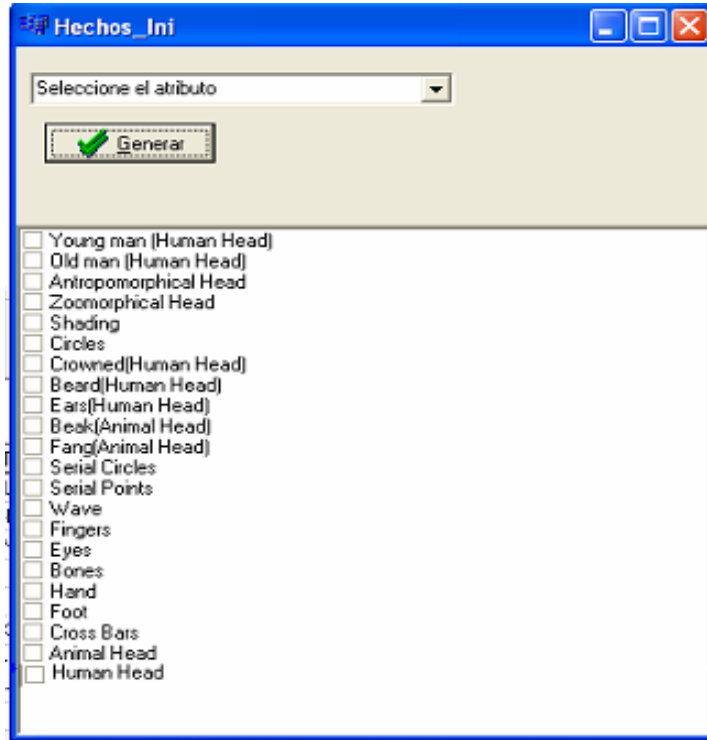
Fig. 4. Dialog for selection of feature set.

Fig. 5. Selection of the initial feature values.

4 Conclusions

We presented the project related to the development of Mayan glyph database, which is the first necessary step in development of computer representation of Mayan script data. The database contains several tables and allows for various predefined queries. The main idea of the project is the development of the system that would allow managing Mayan script data both for specialists and for persons without any previous knowledge of Maya. The system includes structural description of glyph images and a kind of expert system facilities.

We hope that the experts in the field will find this computer tool useful in their daily activities, starting from the usage of the described system that will help them to identify glyphs and reduce the time spent for search and classification of the glyphs.

The developed application will serve also as a didactic tool that helps not only the professional investigators and students, but it will also serve to any person interested in the Mayan writing system.

We plan to implement a determinist expert system based on the structural characteristics of glyph images. This will allow for performing automatic glyph

classification, so that when the database contains the characteristics that correspond to each glyph according to the selected feature set, it will be possible to make reasoning procedures for quick glyph searches.

Fig. 6. Selection of the additional features.
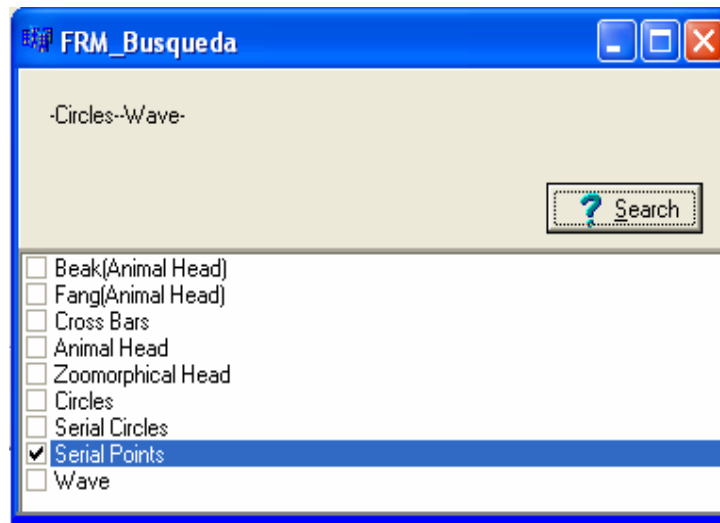
If the glyphs are classified, it will be easier to identify them within documents and Mayan inscriptions. As further investigation, we plan to develop a corpus where the glyphs are represented by their identifiers, for example, Thompson numbers. Note that the majority of existing corpus of Maya is just the digitalized images. Next interesting step in the future research is related to applying statistical methods of corpus linguistics to confirm the results of the deciphering procedures.

Speaking more generally, the presented system can be applied to any set of images that can be assigned a set of structural features. In this perspective, we are considering the implementation not only of the binary features (the feature is present or it is absent), but the features that allow non binary values.

References

1. Foundation for the advancement of Mesoamerican Studies Inc.
   http://www.famsi.org.
2. John Montgomery and Christophe Helmke. Dictionary of Mayan hieroglyphs. Hippocrene Books Inc., 2007.
3. Michael D Coe and Mark Van Stone. Reading the Maya Glyphs. Thames & Hudson, 2001.
4. Inga E. Calvin. Maya Hieroglyphs Study Guide.
   www.famsi.org/mayawriting/calvin/index.html.
5. John Montgomery. How to Read Maya Hieroglyphs. Hippocrene Books Inc., 2003.

6. Alexander Wolfang Voss y Hans Juergen Kremer. Estudio epigráfico sobre las inscripciones jeroglíficas y estudio iconográfico de la Fachada del Palacio de los Estucos en Acanceh Yucatán, México. Instituto de Americanística y Etnología, Universidad de Bonn, Alemania, Noviembre 2000.
7. Yuri V. Knorosov. La antigua escritura de los pueblos de America Central. Biblioteca Obrera, Mexico, 1954.
8. Linda Shele. Notebook for the Maya Hieroglyphic Writing. In: Proc. of Workshop at Texas Austin: Institute of Latin American Studies, University of Texas, 1978.
9. Tatiana Proskourniakoff. Historical data in the inscriptions of Yaxchilan. Part 1. The reign of Shield-Jaguar. Estudios de Cultura Maya (3), Centro de Estudios Mayas, IIFL, UNAM, Mexico, 1963.