The Design and Implementation of AIDA: Ancient Inscription Database and Analytics System

Peter Z. Revesz
University of Nebraska-Lincoln
Lincoln, Nebraska, USA
revesz@cse.unl.edu

M. Parvez Rashid
University of Nebraska-Lincoln
Lincoln, Nebraska, USA
parvezaiub@gmail.com

Yves Tuyishime
University of Nebraska-Lincoln
Lincoln, Nebraska, USA
ytuyishime@unl.edu

ABSTRACT

This paper describes the development of AIDA, the Ancient Inscription Database and Analytics system. The AIDA system currently stores three types of ancient Minoan inscriptions: Linear A, Cretan Hieroglyph and Phaistos Disk inscriptions. In addition, AIDA provides candidate syllabic values and translations of Minoan words and inscriptions into English. The AIDA system allows the users to change these candidate phonetic assignments to the Linear A, Cretan Hieroglyph and Phaistos symbols. Hence the AIDA system provides for various scholars not only a convenient online resource to browse Minoan inscriptions but also provides an analysis tool to explore various options of phonetic assignments and their implications. Such explorations can aid in the decipherment of Minoan inscriptions.

CCS CONCEPTS

• **Information systems → Database design and models**; **Digital libraries and archives**; **Dictionaries**; *Data mining*; *Document representation*; *Specialized information retrieval*.

KEYWORDS

Cretan Hieroglyphs, Data Analytics, Data Mining, Database Design, Linear A, Minoan, User Interface

ACM Reference Format:
Peter Z. Revesz, M. Parvez Rashid, and Yves Tuyishime. 2019. The Design and Implementation of AIDA: Ancient Inscription Database and Analytics System. In *23rd International Database Engineering & Applications Symposium (IDEAS'19), June 10–12, 2019, Athens, Greece.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3331076.3331117

1 INTRODUCTION

The ancient Bronze Age Minoan culture flourished on the island of Crete and some other islands and coastal areas of the Aegean Sea between about 3000 and 1500 BCE [10]. The Minoan language, a Pre-Greek, non-Indo-European language, survives only in ancient inscription in three different types of scripts, namely the Linear A script (about 1500 inscriptions), the Cretan Hieroglyphic script (about 350 inscriptions), and the Phaistos Disk inscription, which has a unique inscription consisting of printed seals for each symbol [11]. There are no widely accepted decipherments of these three Minoan inscriptions, although there are many proposals. One problem with decipherment attempts is that there are too few longer inscriptions in the three different scripts. About 1200 Linear A inscriptions contain only one or two symbols. It would be highly beneficial for a decipherment effort to bring together all three types of inscriptions into a common format. Since Linear A inscriptions are the most common, this would mean in practice the translation of the Cretan Hieroglyph and the Phaistos Disk inscriptions into Linear A. That is one of the goals of our Minoan database system. The basis of the translation to Linear A are two functions. First, a mapping from the Cretan Hieroglyph symbols to the Linear A symbols. Second a mapping from the Phaistos symbols to Cretan Hieroglyph symbols.

We present the AIDA system, short for Ancient Inscription Database and Analytics system, which brings all three types of Minoan inscriptions into the same Linear A format and provides a powerful search capability. The acronym name AIDA is famous from Verdi's opera of the same name, where the Ethiopian princess is called Aida. That name is said to derive from Aita, an ancient Egyptian or other African women's name. It may be also cognate with Finnish äita, which means "mother" in English. In any case, one of the major goals of the AIDA system is to find possible cognates of the Minoan words.

In AIDA, one can enter any Linear A sequence and all the words and the database system will return all the words and inscriptions that contain that sequence including the Cretan Hieroglyph inscriptions and Phaistos Disk blocks whose translations into Linear A contain the search sequence. Similarly, one can search a Cretan Hieroglyph sequence and bring up all three types of inscriptions that contain the equivalent signs. In addition, our system provides the English meaning of a set of words from the lexicon in [16] and translations of texts from [14–17].

The rest of this paper is organized as follows. Section 2 describes all the data sources for our research. Section 3 shows the entity relationship diagram of our database and outlines the main implementation features. Section 4 describes the AIDA system's user interface and some queries. Section 5 outlines the data analytics that the AIDA system is planned to perform. Section 6 discusses related work. Finally, Section 7 gives some conclusions and directions for further research.

2 DATA SOURCES

For the Cretan Hieroglyphic inscriptions we used the book *Corpus Hieroglyphicarum Inscriptionum Cretae*, abbreviated CHIC, by Olivier et al. [12]. For the Linear A inscriptions we used Godart and

IDEAS'19, June 10-12, 2019, Athens, Greece

ACM ISBN 978-1-4503-6249-8/19/06…$15.00
https://doi.org/10.1145/3331076.3331117

Olivier's book *Recueil des inscriptions en Linéaire A* [8], which is commonly abbreviated GORILA by the first letters of the authors and the title. For the Phaistos Disk, we used Evans [7]. These three reference books introduced, respectively, a special numbering of the Cretan Hieroglyph, Linear A and Phaistos Disk symbols. The CHIC book also gave a numbering of the Cretan Hieroglyphic inscriptions. Evans [7] called the two sides of the Phaistos Disk, sides A and B and gave a numbering of the blocks on side A from A1 in the inside to A30 on the outside and on side B from B1 in the inside to B31 on the outside.

Cretan Hieroglyphs and Linear A are just two of about ten different scripts that belong to the *Cretan Script Family*, whose development was studied using bioinformatics phylogenetic algorithms in Revesz [13]. The discovery of the Cretan Script Family played an essential role in the decipherment of the Phaistos Disk [15], Cretan Hieroglyphic inscriptions [17] and Linear A [16]. All these decipherments were based on one-to-one mappings between pairs of scripts within the Cretan Script Family. When a script with known phonetic values is mapped to a script with unknown phonetic values, then the phonetic values of the former script also can be mapped, at least tentatively, to the symbols of the latter script.

Revesz also gave one-to-one mappings from the Phaistos Disk symbols and to the Cretan Hieroglyphs [13] and from the Cretan Hieroglyphs to the Linear A symbols [17]. These mappings enable the transliteration from any of the three types of Minoan inscriptions into the other two types.

3 DATABASE DESIGN AND IMPLEMENTATION

Our entity-relationship diagram is shown in Figure 1. The entity relationship diagram contains a relation for the Phaistos Disk symbols (PD-Symbol), the Cretan Hieroglyph symbols (CH-Symbol) and the Linear A symbols (LA-Symbol). These three sets of symbols are indexed, respectively, by the identification numbers given by Evans [7], CHIC [12], and GORILA [8]. We also have relations that store the Phaistos Disk block numerical sequences (PD-Block), the Cretan Hieroglyph number sequences (CH-Inscriptions), and the Linear A words (Lin-A-Lexicon). Between any type of inscriptions and the corresponding type of symbols, there is a many-to-many containment relation. Therefore, there are three containment relations: Contains-PD, Contains-CH and Contains-LA. Finally, relation Lin-A-inscriptions stores the translated Linear A inscriptions by a number sequence and a meaning. There is a many-to-many relationship between the Lin-A-Lexicon relation and the Lin-A-Inscriptions relation. For each Lin-A-Lexicon tuple we store the Linear A word's number sequence as well as its meaning, which is an English word or phrase. We indicate one-to-one relationships by arrows and the number 1 on the links between the entity sets and the relationship set. Similarly, we also indicate by the symbols *N* and *M* on the links the many-to-many relationships.

For the implementation, we used the MYSQL database system for storing and retrieving data. We built the system interface, which will be described in more detail in Section 4.1, using Boostrap V4.3.1, HTML and CSS. We are running a PHP script to handle the input from the user interface and provide output to the users. The whole system is hosted at the following University of Nebraska-Lincoln server: *https://cse.unl.edu/~revesz/aida.php*.

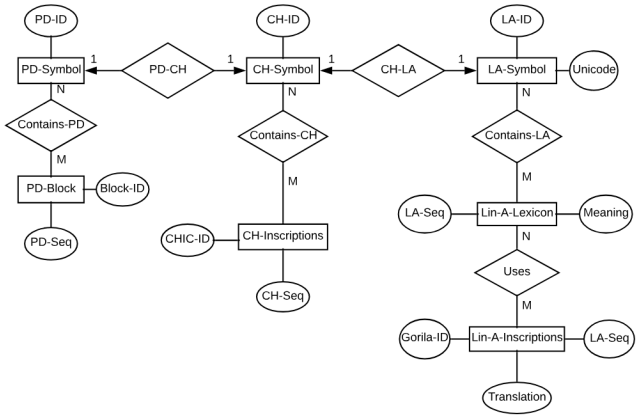

Figure 1: The entity-relationship diagram.

4 THE USER INTERFACE AND QUERIES

Next we describe the AIDA system's user interface in Section 4.1. After that, the following three sections present different types of queries. In particular, Section 4.2 presents Linear A queries, Section 4.3 presents Cretan Hieroglyph queries, and Section 4.4 presents English word queries.

4.1 The User Interface

Figure 2 shows the user interface of the AIDA system. The top line of the user interface contains some clickable choices regarding various information options about the AIDA system, including a brief user's manual that describes how to use the system. The next three lines of the user interface shows three prompt boxes. The user can select any of these three prompt boxes to enter a query. The first prompt box allows the user to enter a Linear A number sequence. The second prompt box allows the user to enter a Cretan Hieroglyph number sequence. The third prompt box allows the user to enter an English keyword. In case the user knows the actual symbol sequence but forgot the associated numbers, the bottom of the AIDA user interface shows a matrix of Linear A symbols. Below each Linear A symbol, its identification number is given based on the GORILA book [8].

4.2 Linear A Queries

By Linear A queries we mean queries that search for the occurrences of various substrings in the Minoan lexicon and the Minoan inscriptions stored in the AIDA system. As an example of a Linear A query, we use the sequence 57-7-67. Given that number sequence, the system returns the answer shown in Figure 3. We see that it is used in three different Linear A inscriptions. For these inscriptions the entire Linear A number sequences and the GORILA identification strings are returned. After the GORILA identification string we also list in parentheses the GORILA volume number and page number separated by a slash where the inscription is described.

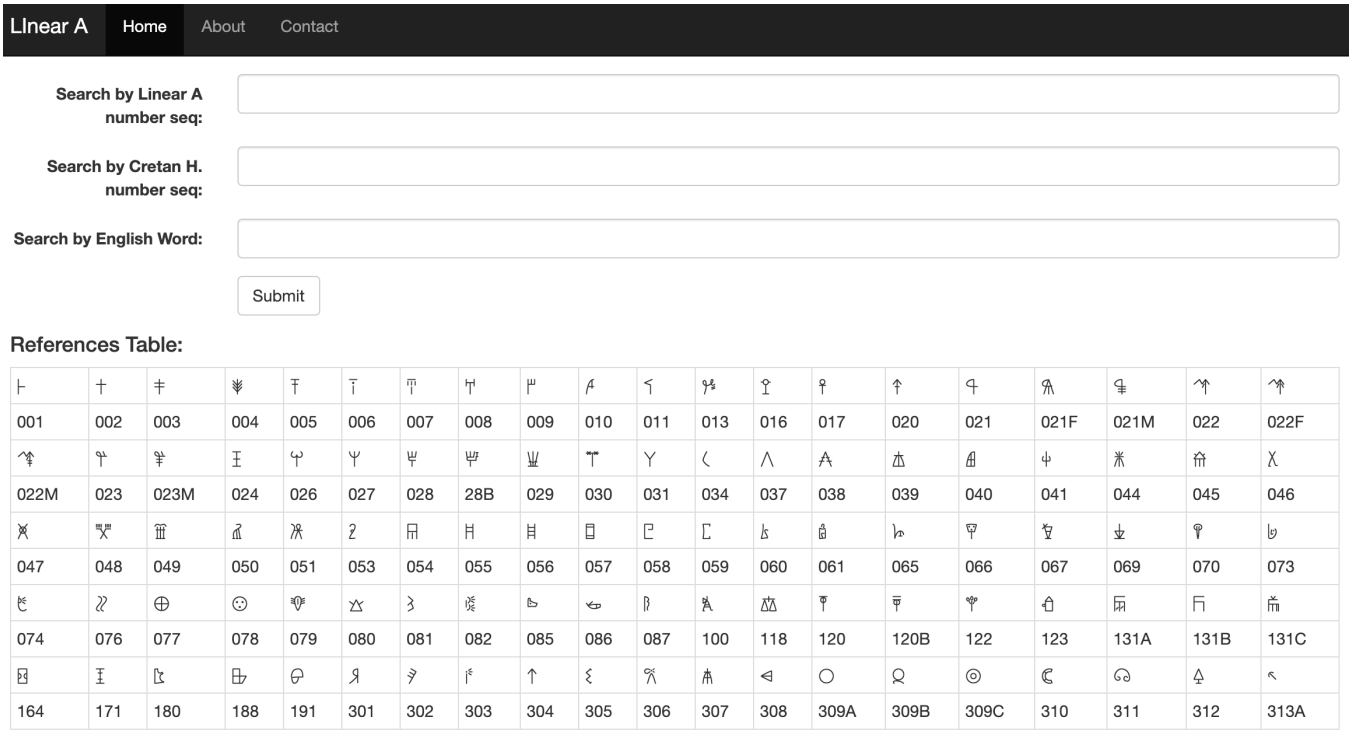

Figure 2: The AIDA user interface.

In addition, the sequence 57-7-67 also occurs in several Linear A lexicon words. One of the lexicon words means "star" while other lexicon words mean "moon." It appears that in the Minoan language the word for "moon" is expressed as either the compound "star+queen" or "star+head", that is, the moon was viewed as the queen or the chief of the stars.

The AIDA system also returns in the last column the syllabic transliteration of the Linear A word for star. The syllabic values are based on Table 12 in [16]. Figure 3 shows that the syllabic value for "star" is *ke-es-ki*. The syllabic values of the Linear A symbols can be updated by the users, which would allow some experimentation. However, any change of syllabic value of a symbol needs to be carefully investigated for its implications. The AIDA system is designed to facilitate such an investigation because the users can retrieve all the words and previous translations that may contain a particular symbol and then see the effect of any change.

The AIDA system also displays in the third and fourth column the putative cognates and the languages in which those cognates occur, respectively. For example, the word *kiška* is a Selkup word that also means "star" in that language. Note the phonetic similarity between *ke-es-ki*, which was likely pronounced as *keski* and the Selkup word *kiška*. The phonetic similarities and the same meaning suggest that they are cognate words. Other possible cognate words retrieved by the AIDA system are *χus* in Khanty, *kōńs* in Mansi and *kušku* in Hattic, all meaning "star."

4.3 Cretan Hieroglyph Queries

Similar to Linear A queries, a Cretan Hieroglyph query retrieves all the Minoan inscriptions that contain a particular Cretan Hieroglyph sequence of its Phaistos Disk or Linear A equivalent sequences. As an example of a Cretan Hieroglyph query, we used the sequence 25-04-03 as shown in Figure 4.

The AIDA system gave an output table where the first column shows the equivalent Linear A sequences of two Minoan inscriptions. The first inscription is a block of the Phaistos Disk, namely block B3. Normally under the CHIC column we would have the Cretan Hieroglyphic inscription identification number from [12], which ranges from #1 to #331. However, there are a few inscriptions that can be considered Cretan Hieroglyph inscriptions, although they do not appear in [12]. One of these inscription is the Arkalochori Axe inscription, which we added to the database as the Cretan Hieroglyphic inscription CHIC #332. The AIDA system was able to bring these two inscriptions with different scripts together and show their relationship. The existence of the common subsequence, which in Linear A would be the following number sequence: 004-712-028, according to the numbering of the Linear A symbols in [8]. The common subsequence implies that it is likely some suffix when the inscriptions are both read from left to right. In a similar manner, a user may find all the occurrences of other candidate prefixes and suffixes. The prefix or suffix nature of the sequences would be strongly supported by their multiple occurrences at the beginning or the end of short inscriptions or the blocks within larger inscriptions such as the Phaistos Disk.

LINEAR A SEQUENCE	Meaning	GORILA	CHIC	PD BLOCK
8-59-28-301-54-57-57-7-67-57-31-31-60-28-39-6-80-41-26-4-59-6-60-4-10-37-55-28-1	All cave spirits: Moon rise IMP big! Cave spirit mother	IO Za 2 (5/19)		
55-56-38-57-7-67-4-4-39-29-27-67-13-28-57-31-10-6-77-6-4-28-51	star [and] Moon ancestor gleam. Blow-V. 3rd SG queen cloud old ancestor	PK Za 8 (4/26)		
57-7-67-4-4-39-29-27	Moon ancestor gleam	PK Za 15 (4/41)		

Meaning	Linear A	Cognates	Language	Syllabic Value
moon	57-7-67-648	cf. star + queen > Kasku	Hattic	
moon	57-7-67-57-31-31-60-13	cf. star + head > Moon		
star	57-7-67	kiška	Selkup	ke-es-ki
star	57-7-67	χus	Khanty	ke-es-ki
star	57-7-67	kōńš	Mansi	ke-es-ki
star	57-7-67	húgy	Hungarian	ke-es-ki
star	57-7-67	kušku	Hattic	ke-es-ki

Figure 3: The result of querying the Linear A sequence 57-7-67.

Linear A	CH or PD	GORILA	CHIC	PD BLOCK
648-017-004-712-028	07-23-35-06-02			B3
031-041-304-004-712-028-029-010-028-086-044-002-712-031-028	27-31-50-25-04-03-66-60-03-40-55-70-04-27-03		332	

Figure 4: The result of querying the Cretan Hieroglyph sequence 25-04-03.

4.4 Word Queries

A word query simply retrieves all the lexicon items and translated texts where some English language keyword appears. The English language keyword can be any word in the English language. If it is not found in the lexicon or the translations, then the AIDA system returns the message "not found." As an example of a word query, we used AIDA to look up all the items that contain the word "light" as shown in Figure 5 and the word "moon" as shown in Figure 6.

As Figure 5 shows, the word "light" occurs not only in the dictionary entry for "light" but also in the dictionary entry for "sunlight." The entry for "light" is associated with two different Linear A number sequences, the first is 8-27 and the second is 8-80, which has syllabic transliterations fe-ne and fe-nu, respectively. These two pronunciations may have been dialectical variations, or they may had slightly different connotations that currently we do not know. However, both of these words seem cognate with other words such as *fény* in Hungarian and *päju* in Sami.

The word for "sunlight" has the Linear A number sequence 302-344-28, syllabic transliteration pj-ai-ku and possible cognate *paike* in the Estonian language, where the word also means "sunlight". More importantly, one can see the possible development from Sami *päju* to Estonian *paike* with a possible suffix -ke at the end of the word.

Figure 6 shows the word query for "moon." As we saw in Section 4.2, in the Minoan language the moon is considered either the queen of stars or the head of stars. Therefore, we see the sequence 57-7-67, which means "star", appear in both definitions of "moon." In addition, the word "moon" appears also in some translated Linear A inscriptions. Finally, Figure 7 shows the word query for "star." It has some overlaps with the previous queries because of the above mentioned reasons.

5 DATA ANALYTICS

The AIDA system can do some simple data analytics. It can count the number of occurrences of any substring. It can also return the most frequent substrings of length k in the inscriptions database, where k is any integer greater than or equal to two. In the future we plan to extend these basic statistics to a more sophisticated analysis where the most frequent substrings are analyzed to check whether they occur preferentially in the beginning, the middle or the end of the inscriptions. This more sophisticated analysis could help determine whether the most frequent substrings are prefixes, word roots, or suffixes, and whether the root words are likely to be nouns

Meaning	Linear A	Cognates	Language	Syllabic Value
light	8-80	fény	Hungarian	fe-nu
light	8-80	bæggjo	Sami	fe-nu
light	8-80	päju	Sami	fe-nu
light	8-27	fény	Hungarian	fe-ne
light	8-27	bæggjo	Sami	fe-ne
light	8-27	päju	Sami	fe-ne
sunlight	302-344-28	paike	Estonian	pj-ai-ku
sunlight	302-344-28	fény	Hungarian	pj-ai-ku
sunlight	302-344-28	fehér	Hungarian	pj-ai-ku

Meaning	Linear A	GORILA ID
[Let the] cloud come, [the] Dan [river] flow, old Tamuz bring heat, shine sunlight	41-41-17-363-310-1-81-73-363-16-73-47-6-60-8-54-39-4-58-45-344-344-28	KN Zf 31 (4/155)

Figure 5: The result of querying the word "light".

Meaning	Linear A	Cognates	Language	Syllabic Value
moon	57-7-67-648	cf. star + queen > Kasku	Hattic	
moon	57-7-67-57-31-31-60-13	cf. star + head > Moon		

Meaning	Linear A	GORILA ID
All cave spirits: Moon rise IMP big! Cave spirit mother	8-59-28-301-54-57-57-7-67-57-31-31-60-28-39-6-80-41-26-4-59-6-60-4-10-37-55-28-1	IO Za 2 (5/19)
All cave spirits, all stars [and the] shiny queen [Moon] cloud-NOUN-PREP run high!	8-59-28-301-54-57-8-7-67-4-41-60-13-8-A363-10-6-26-77-57-41-8-3-51-3-57-57-3-16	PK Za 12 (4/38)
star [and] Moon ancestor gleam. Blow-V. 3rd SG queen cloud old ancestor	55-56-38-57-7-67-4-4-39-29-27-67-13-28-57-31-10-6-77-6-4-28-51	PK Za 8 (4/26)
Moon ancestor gleam	57-7-67-4-4-39-29-27	PK Za 15 (4/41)

Figure 6: The result of querying the word "moon".

or verbs. The AIDA system also could help discover relationships among various scripts, strengthening recent work that shows that Near Eastern scripts have spread both to the west and to the east [4].

6 RELATED WORK

Currently, there is no other online Minoan inscription database system available for public use. However, there is a Linear B inscription database system called the DAMOS system, which is an abbreviation for *Database of Mycenaean at Oslo* [1]. The Linear B script was a successor of the Linear A script [11]. Linear B was the earliest form of Greek writing that is generally agreed to have been deciphered correctly in 1953 by M. Ventris and J. Chandwick [2, 19].

While not a database system, J. Younger's website at the University of Kansas, *http://www.people.ku.edu/ jyounger/LinearA/*, is a frequently consulted online resource for Linear A. It provides an online table of Linear A words with cross references, called "supports" on the website, to all the inscriptions in which the word occurs. Since this website is not a database system, it is not possible to look up in which inscriptions a word occurs by using a simple query. Instead a user needs to manually browse a list of Linear A inscriptions, which are provided on separate webpages, one for the

Meaning	Linear A	Cognates	Language	Syllabic Value
all stars	8-7-67-4	cf. all, star		fe-es-ki-se
chief star	8-7-67	cf. head, star		fe-es-ki
star	57-7-67	kiška	Selkup	ke-es-ki
star	57-7-67	χus	Khanty	ke-es-ki
star	57-7-67	kõñš	Mansi	ke-es-ki
star	57-7-67	húgy	Hungarian	ke-es-ki
star	57-7-67	kušku	Hattic	ke-es-ki
star	56-38	csillag	Hungarian	za-la

Meaning	Linear A	GORILA ID
[Sun] shine-IMP and [stars] gleam-IMP down happy love-ACC every day	8-27-24-27-7-301-39-44-24-57-59-53-28-453-23-8-57-37	KN Zf 13 (4/153)
All cave spirits, all stars [and the] shiny queen [Moon] cloud-NOUN-PREP run high!	8-59-28-301-54-57-8-7-67-4-41-60-13-8-A363-10-6-26-77-57-41-8-3-51-3-57-57-3-16	PK Za 12 (4/38)
All cave spirit-INSTR. chief star ancestor gleam down love fa ko fa j chief queen cloud-POSS-PREP rise IMP big out high!	8-59-28-301-54-38-8-7-67-4-4-1-39-4-53-8-70-8-363-8-31-31-60-13-10-6-26-77-6-34-28-99-6-73-6-41-26-28-6-57-3-16	PK Za 11 (4/34)
star [and] Moon ancestor gleam. Blow-V. 3rd SG queen cloud old ancestor	55-56-38-57-7-67-4-4-39-29-27-67-13-28-57-31-10-6-77-6-4-28-51	PK Za 8 (4/26)

Figure 7: The result of querying the word "star".

Haghia Triada inscriptions, another for the Knossos inscriptions, and so on at each separate location.

7 CONCLUSIONS AND FUTURE WORK

The development of the AIDA system is challenging because it requires knowledge of the important database system design principles as well as a knowledge of Minoan inscriptions and the basic concepts of comparative linguistics. These three areas of knowledge are uniquely brought together in our AIDA system. The AIDA system has a potential to be a widely used resource for many scholars in the humanities in the fields of classics, history and linguistics. As a future work, we hope to extend the system with other ancient languages, such as Sumerian [5, 18], Elamite [6], and the Indus Valley Script [3, 20]. As our database grows, we also investigate the possibility of using ElasticSearch [9] to make queries more efficient.

REFERENCES

[1] F. Aurora, A. Nesøen, D. Nedić, H. Løken, and A. Bersi. DAMOS - Database of Mycenaean at Oslo, 2018.

[2] J. Chadwick. *The Decipherment of Linear B.* Cambridge University Press, 1958.

[3] S. Daggumati and P. Z. Revesz. Data mining ancient script image data using convolutional neural networks. In *Proceedings of the 22nd International Database Engineering and Applications Symposium*, pages 267–272. ACM, 2018.

[4] S. Daggumati and P. Z. Revesz. Data mining ancient scripts to investigate their relationships and origins. In *Proceedings of the 23rd International Database Engineering and Applications Symposium*, 2019.

[5] C. Elisabeth and D. Caspers. Sumer, coastal Arabia and the Indus Valley in protoliterate and early dynastic eras: Supporting evidence for a cultural linkage. *Journal of the Economic and Social History of the Orient/Journal de l'histoire economique et sociale de l'Orient*, pages 121–135, 1979.

[6] R. K. Englund. The Proto-Elamite script. In P. T. Daniels and W. Bright, editors, *The World's Writing Systems*, pages 160–164. Oxford University Press, 1996.

[7] A. J. Evans. *Scripta Minoa: The Written Documents of Minoa Crete with Special Reference to the Archives of Knossos.* Classic Books, 1909.

[8] L. Godart and J.-P. Olivier. *Recueil des inscriptions en Linéaire A.* Number 21 in Études Crétoises. De Boccard, 1976.

[9] C. Gormley and Z. Tong. *Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine.* O'Reilly Media, Inc., 2015.

[10] N. Marinatos. *Minoan Kingship and the Solar Goddess: A Near Eastern Koine.* University of Illinois Press, 2010.

[11] J.-P. Olivier. Cretan writing in the second millennium BC. *World Archaeology*, 17(3):377–389, 1986.

[12] J.-P. Olivier, L. Godart, and J.-C. Poursat. *Corpus Hieroglyphicarum Inscriptionum Cretae*, volume 31 of *Études Crétoises.* De Boccard, 1996.

[13] P. Z. Revesz. Bioinformatics evolutionary tree algorithms reveal the history of the Cretan Script Family. *International Journal of Applied Mathematics and Informatics*, 10:67–76, 2016.

[14] P. Z. Revesz. A computer-aided translation of the Cretan Hieroglyph script. *International Journal of Signal Processing*, 1:127–133, 2016.

[15] P. Z. Revesz. A computer-aided translation of the Phaistos Disk. *International Journal of Computers*, 10:94–100, 2016.

[16] P. Z. Revesz. Establishing the West-Ugric language family with Minoan, Hattic and Hungarian by a decipherment of Linear A. *WSEAS Transactions on Information Science and Applications*, 14:306–335, 2017.

[17] P. Z. Revesz. A translation of the Arkalochori Axe and the Malia Altar Stone. *WSEAS Transactions on Information Science and Applications*, 14(1):124–133, 2017.

[18] P. Z. Revesz. Sumerian contains Dravidian and Uralic substrates associated with the Emegir and Emesal dialects. *WSEAS Transactions on Information Science and Applications*, 16(1):8–30, 2019.

[19] M. Ventris and J. Chadwick. *Documents in Mycenaean Greek.* Cambridge University Press, 2nd edition, 1973.

[20] B. K. Wells and A. Fuls. *The Archaeology and Epigraphy of Indus Writing.* Archaeopress, 2015.