Summer 7-24-2019

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

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

THE DESIGN AND IMPLEMENTATION OF AIDA: ANCIENT INSCRIPTION DATABASE AND ANALYTICS SYSTEM

by

M Parvez Rashid

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfilment of Requirements

For the Degree of Master of Science

Major: Computer Science

Under the Supervision of Professor Peter Z. Revesz

Lincoln, Nebraska

July, 2019

THE DESIGN AND IMPLEMENTATION OF AIDA: ANCIENT INSCRIPTION DATABASE AND ANALYTICS SYSTEM

M Parvez Rashid, M.S.

University of Nebraska, 2019

Adviser: Peter Z. Revesz

AIDA, the Ancient Inscription Database and Analytic system can be used to translate and analyze ancient Minoan language. 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.

DEDICATION

To my parent

ACKNOWLEDGMENTS

I would like to express my gratitude and appreciation to my thesis advisor Prof. Peter Revesz of the department of computer science at the University of Nebraska-Lincoln. Prof. Revesz was always ready for my questions about my research or writing. Not only his great ideas steered me in the right direction, but without his support, the completion of this work was not possible.

I would like to acknowledge Prof. Hongfeng Yu and Prof. Mohammad Hasan from the Computer science department at the University of Nebraska-Lincoln as the committee members of this thesis, and I am gratefully indebted to them for their precious comments on this thesis. I would also like to acknowledge Yves Tuyishime for his support.

Author

M Parvez Rashid

Table of Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Minoan Language

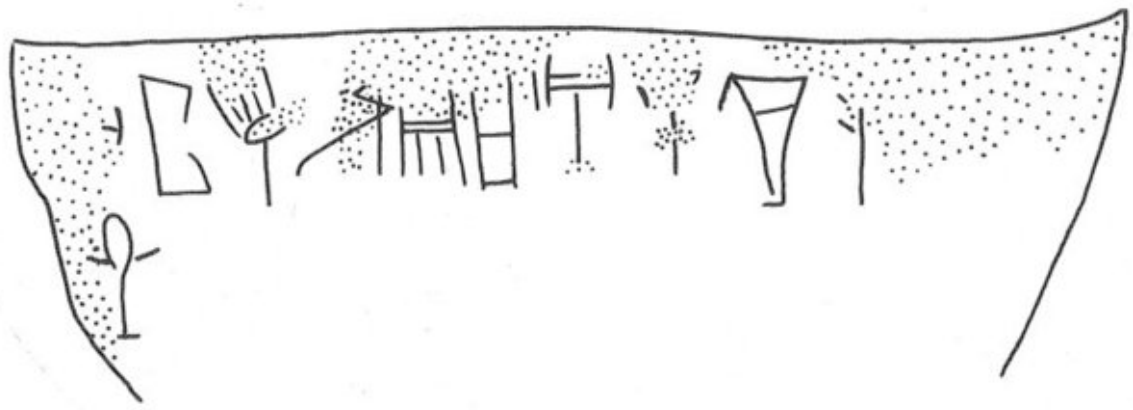
The Minoan language was the language of the ancient Minoan civilization of Crete. During the Bronze Age Minoan culture flourished on the island of Crete and some other islands and coastal areas of the Aegean Sea between about 3000 and 1500 BCE [35]. The Minoan language, a Pre-Greek, non-Indo-European 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 is a unique inscription consisting of printed seals for each symbol [38]. Despite the available evidence and inscriptions, the Minoan language is unknown and unclassified at this present date.

1.1.1 Linear A inscription

Inscriptions written by Minoan civilization using Cretan hieroglyphs and later in the Linear A syllabary are still undeciphered except the fact that Linear A only partly deciphered as they are fairly legible by comparison with Linear B. Linear A is the main Minoan script. It was discovered in Knossos by archaeologist Arthur Evans [17]. Evans named the script "Linear" as its characters consisted of lines inscribed in clay. In contrast,

Cretan hieroglyphs had more pictographic characters, which were used during the same period. Linear A flourished during the Middle Minoan Period, specifically from $1625 - 1450$ BC [38]. It is a possible descendent of Cretan hieroglyphs and the ancestor of Linear B. Despite being the main Minoan script, Linear A remained undeciphered for over a century. In the Figure 1.1, a Linear A inscription is presented. In Godart and Oliver [22] book it is included with a clear drawing of the original inscription.

(a) Linear A original inscription

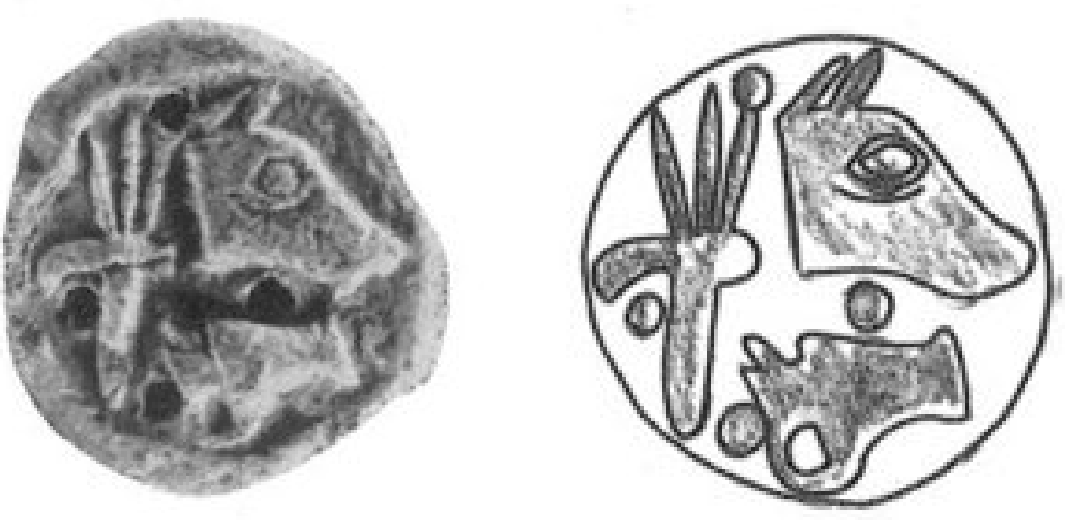


(b) Recreated drawing of the inscription

Figure 1.1: Linear A Inscription on a clay pot (Source: Godart and Olivier [22])

1.1.2 Cretan Hieroglyphs

The Cretan Hieroglyphs script was used on the Island of Crete between $2000 - 1650$ BC as a writing system. Though it predates Linear A by about a century, both writing systems continued to be used in parallel for most of their history. Most of the symbols found in the Cretan Hieroglyphs writing were found in the form of clay documents. Some of the inscriptions were incised on the still moist clay with a sharp stylus and some were also carved on minute stone seals [2]. It is a belief that, like most other hieroglyphic writing systems, for example Egyptian, Hittite, even the Mayan, the Cretan Hieroglyphs script was developed in the basis of culture, rituals, and indigenous practices. Figure 1.1 shows a sample Cretan Hieroglyph inscription.



(a) CH original inscription (b) Drawing of the original inscription

Figure 1.2: Cretan Hieroglyphs Inscription (Source: CHIC by J.P. Olivier, L. Godart, and J.C. Poursat [40]))

1.1.3 Phaistos Disk

The Phaistos Disk is a 4000 year old clay disk, discovered in 1908 by the Italian archaeologist Luigi Pernier in the Minoan palace-site of Phaistos on the Greek island of Crete.

The Phaistos Disk contains 241 tokens, comprising 45 distinct signs, which were made by pressing hieroglyphic seals in a spiral arrangement into a disc of soft clay. The diameter of the disk varies between 15.8 cm to 16.5 cm and its thickness varies between 1.6 to 2.1 cm. The variation in size indicates that it was handmade. The imperfection of the spiral lines also indicate it was hand drawn. The symbols are grouped between 2 and 7 symbols by vertical lines. The direction of reading the symbols and their linguistic meaning is still a mystery. Figure 1.3 shows the two faces of the Phaistos Disk.

(a) Phaistos Disk Face A (b) Phaistos Disk Face B

Figure 1.3: The Phaistos Disk made of fired clay found in Crete [37]

1.2 Problem Statement

There are no widely accepted decipherments of the Minoan inscriptions in Cretan Hieroglyphs, Linear A and Phaistos Disk although there are many proposals. Despite of having many inscriptions from Knossos and other places, the Minoan language remains the only extensive writing of the ancient European civilization which can not yet be deciphered. It

is mostly because the inscriptions are written in very abbreviated form and the literature is not extensive. Another 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 is an extremely difficult task to decipher literature that is only found in a limited number written on clay tablet heavily damaged and buried for centuries. The only hope to decipher the Minoan language is to co-ordinate and review the inscriptions carefully and find the relationship between the sentences and words in those three different inscriptions.

1.3 Objective

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 Disk symbols to Cretan Hieroglyph symbols.

1.4 Contribution

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 any English word 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 [57] and translations of texts from [55, 54, 57, 58].

1.5 Outline of the Thesis

We start with a general overview of the thesis topics, the first chapter briefly explains the objectives, problems, the final goal, and potential contributions of the study. Chapter 2 describes some related work. To our knowledge AIDA is the first attempt to make a database for storing the three different types of inscriptions of the in Minoan language. There are some other web-pages that present some information regarding Minoan inscriptions, but they can not be queried and easily used for data analysis and finding relations between three different inscriptions types. Chapter 3 presents the data sources for the inscriptions that are used for this thesis and most importantly the one-to-one relation mappings between pairs of inscriptions and symbols. In Chapter 4 we show the data statistics of AIDA system. In Chapter 5, a brief description of the database structure is given together with the application interface. Chapter 6 reports the tabulated results of this application followed by some discussion. Chapter 7 contains the conclusions, limitations, and some proposed suggestions regarding future work and potential improvements of the AIDA system.

Chapter 2

RELATED WORKS

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* [4]. The Linear B script was a successor of the Linear A script [38]. 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 [7, 67].

Chapter 3

Data Sources

For the Cretan Hieroglyphic inscriptions we used the book *Corpus Hieroglyphicarum Inscriptionum Cretae*, abbreviated CHIC, by Olivier et al. [40]. For the Linear A inscriptions we used Godart and Olivier's book *Recueil des inscriptions en Linéaire A* [22], which is commonly abbreviated GORILA by the first letters of the authors and the title. For the Phaistos Disk, we used Evans [17]. 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 [17] 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 [53]. The discovery of the Cretan Script Family played an essential role in the decipherment of the Phaistos Disk [55], Cretan Hieroglyphic inscriptions [58] and Linear A [57]. 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 [53] and from the Cretan Hieroglyphs to the Linear A symbols [58]. These mappings enable the transliteration from any of the three types of Minoan inscriptions into the other two types.

Cretan Hieroglyph	Linear A	Cretan Hieroglyph	Linear A	Cretan Hieroglyph	Linear A
001	46	033	309	065	7
002	352	034	118	066	29
003	28	035	58	067	3
004	712	036	22	068	24
005	309	037	307	069	76
006	48	038	56	070	2
007	58	039	351	071	314
008	362	040	86	072	67
009	55	041	54	073	77
010	338	042	317	074	78
011	418	043	363	075	78
012	418	044	320	076	164
013	21	045	8	077	311
014	85	046	34	078	326
015	59	047	78	079	51
016	85	048	13	080	732
017	16	049	37	081	309
018	301	050	304	082	407
019	357	051	312	083	409
020	45	052	50	084	57
021	39	053	123	085	80
022	39	054	648	086	180
023	122	055	44	087	315
024	30	056	61	088	87
025	4	057	1	089	81
026	359	058	55	090	60
027	31	059	10	091	37
028	302	060	10	092	26
029	73	061	53	093	37
030	27	062	17	094	38
031	41	063	6	095	362
032	41	064	47	096	321

Figure 3.1: One-to-one mappings from Cretan Hieroglyphic to Linear A (Source: Revesz et al. [58])

1		001		16		045		31		092	
2		003		17		076		32		052	
3		044		18		007		33		019	
4		006		19		027		34		020	
5		002		20		053		35		025	
6		004		21		068		36		024	
7		054		22		055		37		078	
8		009		23		062		38		005	
9		046		24		037		39		023	
10		049		25		040		40		153	
11		048		26		060		41		066	
12		047		27		041		42		039	
13		067		28		010		43		072	
14		034		29		017		44		051	
15		042		30		013		45		069	

(a) PD to CH 1-15 (b) PD to CH 16-31 (c) PD to CH 31-45

Figure 3.2: One-to-one mappings from Phaistos disk to Cretan Hieroglyphic(Source: Revesz et al. [53])

Chapter 4

Data Statistics

AIDA has a total of 661 inscriptions of which 91 are Linear A inscriptions, 509 Cretan Hieroglyph and 61 Phaistos Disk inscriptions. For Linear A we took the inscription sequence from GORILA [22] and for Cretan Hieroglyphs we used the inscriptions available in CHIC [40].

Inscription type	Count
Linear A	91
Phaistos Disk	61
Cretan Hieroglyph	509
Total	661

Table 4.1: Inscription types and counts

For representing the signs with id numbers AIDA has 260 ids for Linear A symbols, 153 for Cretan Hieroglyph and 45 for the Phaistos Disk symbols.

Sign type	Count
Linear A	260
Phaistos Disk	45
Cretan Hieroglyph	153
Total	458

Table 4.2: Sign types and counts

AIDA has 90 English words available for corresponding Linear A symbol sequence. We used these words from Revesz [57].

English Words	90
Different Cognates	209
Other Languages	10

Table 4.3: English to Linear A words with cognates

For translating to Linear A from Cretan Hieroglyph and Phaistos Disk, AIDA has 96 entries for Cretan Hieroglyph to linear A and 45 entries for Phaistos Disk to Cretan Hieroglyph. We collected the correspondence table from Revesz [53].

Correspondence of Alphabets	
Cretan Hieroglyph to Linear A	96
Phaistos Disk to Cretan Hieroglyph	45

Table 4.4: Correspondence of alphabets between three different Minoan writing systems

Chapter 5

Database Design and Implementation

Our entity-relationship diagram is shown in Figure 5.1. 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 [17], CHIC [40], and GORILA [22]. 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 5.1.1, 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.

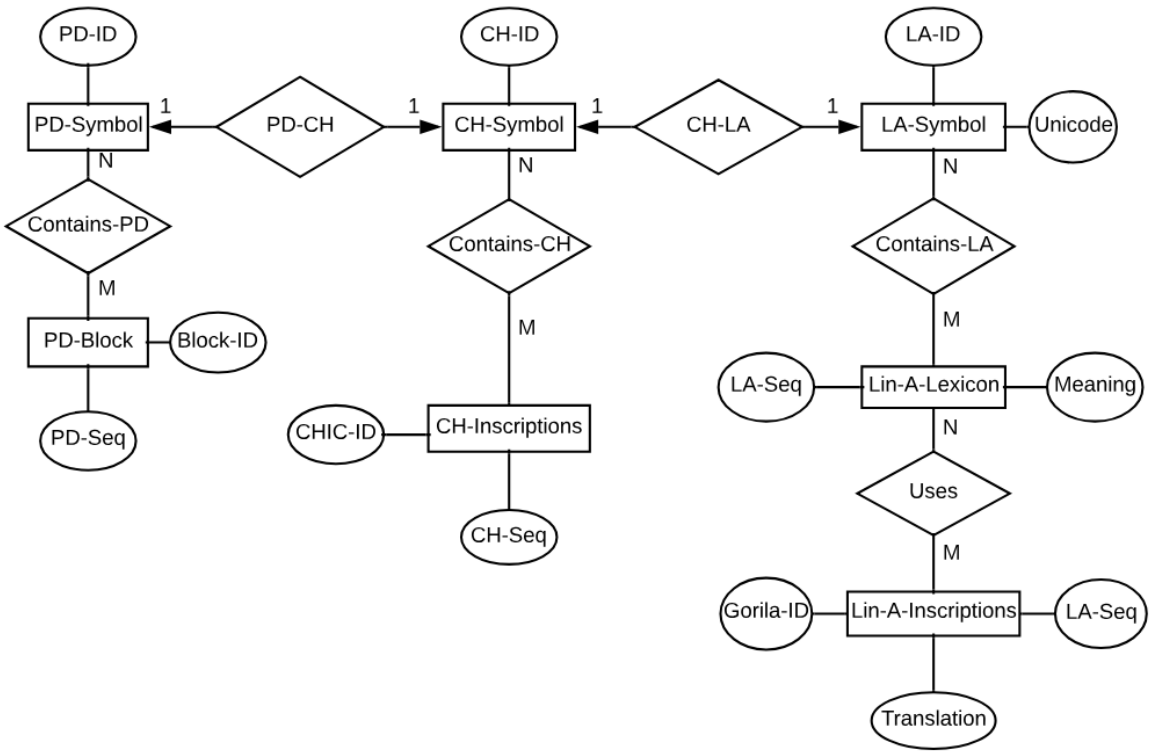


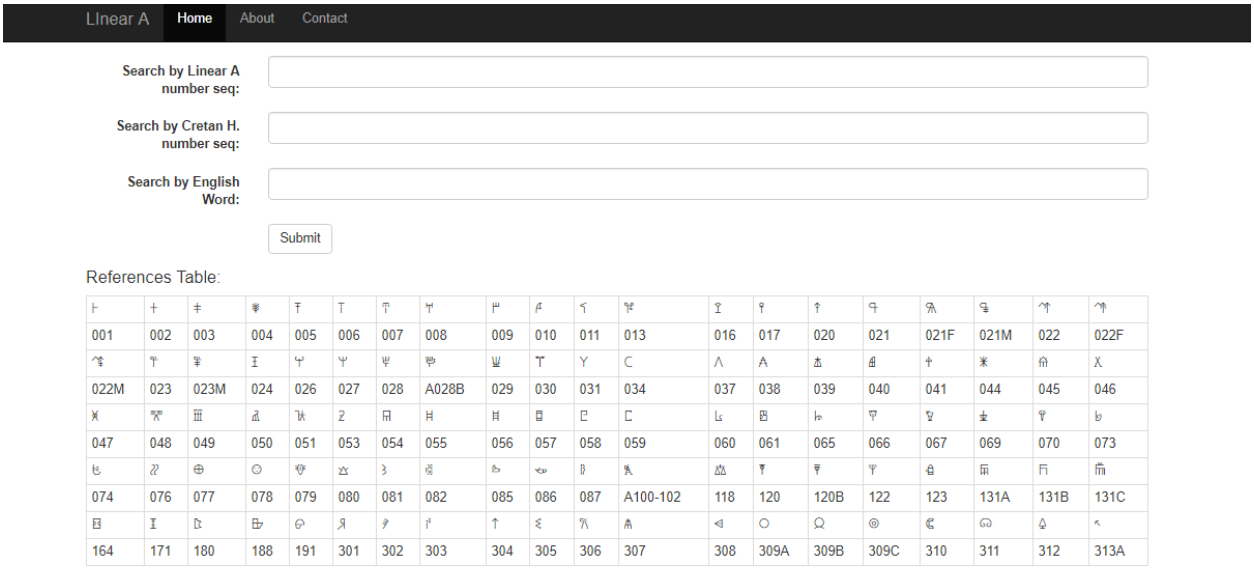
Figure 5.1: The entity-relationship diagram (Source: Revesz et al. [61]).

5.1 The User Interface and Queries

Next we describe the AIDA system's user interface in Section 5.1.1. After that, the following three sections present different types of queries. In particular, Section 5.1.2 presents Linear A queries, Section 5.1.3 presents Cretan Hieroglyph queries, and Section 5.1.4 presents English word queries.

5.1.1 The User Interface

Figure 5.2 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 [22].



001	002	003	004	005	006	007	008	009	010	011	013	016	017	020	021	021F	021M	022	022F
022M	023	023M	024	026	027	028	A028B	029	030	031	034	037	038	039	040	041	044	045	046
047	048	049	050	051	053	054	055	056	057	058	059	060	061	065	066	067	069	070	073
074	076	077	078	079	080	081	082	085	086	087	A100-102	118	120	120B	122	123	131A	131B	131C
164	171	180	188	191	301	302	303	304	305	306	307	308	309A	309B	309C	310	311	312	313A

Figure 5.2: The AIDA user interface.

5.1.2 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 5.3. 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.

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 [57]. Figure 5.3 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".

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

5.1.3 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 5.4.

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 [40], 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 [40]. 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 [22]. 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	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 5.4: The result of querying the Cretan Hieroglyph sequence 25-04-03.

5.1.4 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 5.5 and the word "moon" as shown in Figure 5.6.

As Figure 5.5 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

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 5.5: The result of querying the word ”light”.

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 5.6 shows the word query for “moon”. As we saw in Section 5.1.2, 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 5.7 shows the word query for “star”. It has some overlaps with the previous queries because of the above mentioned reasons.

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 5.6: The result of querying the word ″moon″.

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	kōńš	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 5.7: The result of querying the word ″star″.

Chapter 6

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 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 [12].

Chapter 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 [14, 59], Elamite [16], and the Indus Valley Script [11, 71]. As our database grows, we also investigate the possibility of using ElasticSearch [23] to make queries more efficient.

Bibliography

[1] Abien Fred Agarap. An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification. *arXiv preprint arXiv:1712.03541*, 2017.

[2] Rutger Allan. *ENCYCLOPEDIA OF ANCIENT GREEK LANGUAGE AND LINGUISTICS*. 2015. 1.1.2

[3] Sandhya Arora, Debotosh Bhattacharjee, Mita Nasipuri, Latesh Malik, Mohantapash Kundu, and Dipak Kumar Basu. Performance comparison of SVM and ANN for handwritten Devnagari character recognition. *arXiv preprint arXiv:1006.5902*, 2010.

[4] F. Aurora, A. Nesøen, D. Nedić, H. Løken, and A. Bersi. DAMOS - Database of Mycenaean at Oslo, 2018. 2

[5] Jan G. P. Best and Fred C. Woudhuizen, editors. *Ancient Scripts from Crete and Cyprus*, volume 9. Bill Archive, 1988.

[6] Vaclav Blazek. Elam: a bridge between ancient near east and dravidian india. *Archaeology and Language IV. Language Change and Cultural Transformation. London, New York: Routledge*, pages 48–78, 1999.

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

[8] Dominique Collon. Mesopotamia and the Indus: The evidence of the seals. In *The Indian Ocean in Antiquity*, pages 209–225. The British Museum and Kegan Paul International London/New York, 1996.

[9] Brian Francis Cook. *Greek Inscriptions*, volume 5. University of California Press, 1987.

[10] Shruti Daggumati. Similarity queries on script image databases. In András Benczúr, Bernhard Thalheim, Tomás Horváth, Silvia Chiusano, Tania Cerquitelli, Csaba István Sidló, and Peter Z. Revesz, editors, *New Trends in Databases and Information Systems - ADBIS 2018 Short Papers and Workshops*, pages 391–401. Springer, 2018.

[11] Shruti Daggumati and Peter 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. 7

[12] Shruti Daggumati and Peter Z. Revesz. Data mining ancient scripts to investigate their relationships and origins. In *Proceedings of the 23rd International Database Engineering and Applications Symposium*, 2019. 6

[13] Jacob L Dahl. Complex graphemes in Proto-Elamite. *Cuneiform Digital Library Journal*, 2005(6), 2005.

[14] CL Elisabeth and During 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. 7

[15] Mohamed Elleuch, Najiba Tagougui, and Monji Kherallah. A novel architecture of CNN based on SVM classifier for recognizing Arabic handwritten script. *International Journal of Intelligent Systems Technologies and Applications*, 15(4):323–340, 2016.

[16] Robert K Englund. The Proto-Elamite script. In Peter T. Daniels and William Bright, editors, *The World's Writing Systems*, pages 160–164. Oxford University Press, 1996. 7

[17] A. J. Evans. *Scripta Minoa: The Written Documents of Minoa Crete with Special Reference to the Archives of Knossos*. Classic Books, 1909. 1.1.1, 3, 5

[18] Walter Ashlin Fairservis. *The Harappan civilization and its writing: a model for the decipherment of the Indus script*. Brill, 1992.

[19] Steve Farmer, Richard Sproat, and Michael Witzel. The collapse of the Indus-script thesis: The myth of a literate Harappan civilization. *Electronic Journal of Vedic Studies*, 11(2):19–57, 2016.

[20] Steven R Fischer. *History of Writing*. Reaktion Books, 2004.

[21] Andreas Fuls. Online Archaeoastronomy in Space and Time. http://www.user.tu-berlin.de/fuls/Homepage/, 2018.

[22] L. Godart and J.-P. Olivier. *Recueil des inscriptions en Linéaire A*. Number 21 in Études Crétoises. De Boccard, 1976. (document), 1.1.1, 1.1, 3, 4, 5, 5.1.1, 5.1.3

[23] Clinton Gormley and Zachary Tong. *Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine*. O'Reilly Media, Inc., 2015. 7

[24] Meijun He, Shuye Zhang, Huiyun Mao, and Lianwen Jin. Recognition confidence analysis of handwritten Chinese character with CNN. In *Proceedings of the 13th International Conference on Document Analysis and Recognition*, pages 61–65. IEEE, 2015.

[25] James T. Hooker and John H. Betts. *Linear B: An Introduction*. Bristol Classical Press, Bristol, UK, 1980.

[26] Michael C Howard. *Transnationalism in Ancient and Medieval Societies: The Role of Cross-Border Trade and Travel*. McFarland, 2014.

[27] Max Jaderberg, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Synthetic data and artificial neural networks for natural scene text recognition. *arXiv preprint arXiv:1406.2227*, 2014.

[28] Max Jaderberg, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Reading text in the wild with convolutional neural networks. *International Journal of Computer Vision*, 116(1):1–20, 2016.

[29] Keechul Jung, Kwang In Kim, and Anil K Jain. Text information extraction in images and video: A survey. *Pattern Recognition*, 37(5):977–997, 2004.

[30] Subhash C Kak. A frequency analysis of the Indus script. *Cryptologia*, 12(3):129–143, 1988.

[31] David H Kelley and Bryan Wells. Recent Developments In Understanding The Indus Script. *The Quarterly Review of Archaeology*, 16(1):15–23, 1995.

[32] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations*, 2015.

[33] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[34] Yann LeCun, Corinna Cortes, and Chris Burges. MNIST handwritten digit database. http://yann.lecun.com/exdb/mnist/, 1998.

[35] N. Marinatos. *Minoan Kingship and the Solar Goddess: A Near Eastern Koine*. University of Illinois Press, 2010. 1.1

[36] David W McAlpin. Proto-Elamo-Dravidian: The evidence and its implications. *Transactions of the American Philosophical Society*, 71(3):1–155, 1981.

[37] J. P. Olivier. *Le Disque de Phaistos*. Bulletin de Correspondance Hellnique, 1975. (document), 1.3

[38] J-P Olivier. Cretan writing in the second millennium BC. *World Archaeology*, 17(3):377–389, 1986. 1.1, 1.1.1, 2

[39] J. P. Olivier. *Cretan Writing in the Second Millennium B.C.* World Archaeology, 1986.

[40] J.-P. Olivier, L. Godart, and J.-C. Poursat. *Corpus Hieroglyphicarum Inscriptionum Cretae*, volume 31 of *Études Crétoises*. De Boccard, 1996. (document), 1.2, 3, 4, 5, 5.1.3

[41] Asko Parpola. The Indus script: A challenging puzzle. *World Archaeology*, 17(3):399–419, 1986.

[42] Asko Parpola. The Indus Script. In Peter T. Daniels and William Bright, editors, *The World's Writing Systems*, pages 165–171. Oxford University Press, 1996.

[43] Asko Parpola. Study of the Indus script. In *Proceedings of the International Conference of Eastern Studies*, volume 50, pages 28–66, 2005.

[44] Asko Parpola. *Deciphering the Indus script*. Cambridge University Press, 2009.

[45] Simo Parpola. *Etymological Dictionary of the Sumerian Language*, volume 1 and 2. Foundations for Finnish Assyrological Research, Helsinki, Finnland, 2016.

[46] Rajesh PN Rao, Nisha Yadav, Mayank N Vahia, Hrishikesh Joglekar, R Adhikari, and Iravatham Mahadevan. Entropic evidence for linguistic structure in the Indus script. *Science*, 324(5931):1165–1165, 2009.

[47] Rajesh PN Rao, Nisha Yadav, Mayank N Vahia, Hrishikesh Joglekar, R Adhikari, and Iravatham Mahadevan. A Markov model of the Indus Script. *Proceedings of the National Academy of Sciences*, 106(33):13685–13690, 2009.

[48] Shikaripur Ranganatha Rao. *The Decipherment of the Indus Script*. Asia Publishing House, 1982.

[49] Colin Renfrew. Archaeology, genetics and linguistic diversity. *Man*, pages 445–478, 1992.

[50] P. Z. Revesz. Spatio-temporal data mining of major european river and mountain names reveals their near eastern and african origins. In *22nd European Conference on Advances in Databases and Information Systems*, pages 20–32. Springer LNCS, 2018.

[51] Peter Z. Revesz. *Introduction to Databases: From Biological to Spatio-Temporal*. Springer, 2010.

[52] Peter Z. Revesz. An algorithm for constructing hypothetical evolutionary trees using common mutations similarity matrices. In *Proc. 4th ACM International Conference on Bioinformatics and Computational Biology (ACM BCB)*, pages 731–734, 2013.

[53] Peter 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. (document), 3, 3.2, 4

[54] Peter Z. Revesz. A computer-aided translation of the Cretan Hieroglyph script. *International Journal of Signal Processing*, 1:127–133, 2016. 1.4

[55] Peter Z. Revesz. A computer-aided translation of the Phaistos Disk. *International Journal of Computers*, 10:94–100, 2016. 1.4, 3

[56] Peter Z. Revesz. A mitochondrial DNA-based model of the spread of human populations. *International Journal of Biology and Biomedical Engineering*, 10:124–133, 2016.

[57] Peter 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. 1.4, 3, 4, 5.1.2

[58] Peter 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. (document), 1.4, 3, 3.1

[59] Peter 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. 7

[60] Peter Z. Revesz and Zhiqiang Li. Incremental phylogenetics by repeated insertions: An evolutionary tree algorithm. *International Journal of Biology and Biomedical Engineering*, 10(1):148–158, 2016.

[61] Peter Z. Revesz, M. Parvez Rashid, and Yves Tuyishime. The design and implementation of AIDA: Ancient Inscription Database and Analytics system. In *Proceedings of the 23rd International Database Engineering and Applications Symposium*, 2019. (document), 5.1

[62] Richard Salomon. *Deciphering the Indus Script*. JSTOR, 1996.

[63] Richard Salomon. *Indian Epigraphy: A Guide to the Study of Inscriptions in Sanskrit, Prakrit, and the Other Indo-Aryan Languages*. Oxford University Press, 1998.

[64] Richard G Salomon. Brahmi and kharoshthi. *The world's writing systems*, pages 373–383, 1996.

[65] Upinder Singh. *A history of ancient and early medieval India: from the Stone Age to the 12th century*. Pearson Education India, 2008.

[66] Yichuan Tang. Deep learning using linear support vector machines. *arXiv preprint arXiv:1306.0239*, 2013.

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

[68] Bryan Wells. *An introduction to Indus writing*. University of Calgary, 1998.

[69] Bryan Wells and Andreas Fuls. Online Indus Writing Database. http://caddy.igg.tu-berlin.de/indus/welcome.htm, 2017.

[70] Bryan K Wells. *Epigraphic Approaches to Indus Writing*. Oxbow Books, 2011.

[71] Bryan K Wells and Andreas Fuls. *The Archaeology and Epigraphy of Indus Writing*. Archaeopress, 2015. 7

[72] Wikipedia. Aida (given name).

[73] Weixin Yang, Lianwen Jin, and Manfei Liu. Chinese character-level writer identification using path signature feature, DropStroke and deep CNN. In *Proceedings of the 13th International Conference on Document Analysis and Recognition*, pages 546–550. IEEE, 2015.

[74] Margot Lisa-Jing Yann and Yichuan Tang. Learning deep convolutional neural networks for X-ray protein crystallization image analysis. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 1373–1379. AAAI Press, 2016.

[75] J. G. Younger. Linear a texts and inscriptions in phonetic transcription.

[76] Arlene R Zide and Kamil V Zvelebil. *The Soviet Decipherment of the Indus Valley Script: Translation and Critique*, volume 156. Walter de Gruyter, 1976.