othe elde and a contracted of he elde and a contracted at the fille and a contracted at the fille and a contracted at the section of the sect

Searching Four-Millennia-Old Digitized Documents: A Text Retrieval System for Egyptologists

Estíbaliz Iglesias-Franjo, Jesús Vilares

Language in the Information Society Group Universidade da Coruña <u>www.grupolys.org</u>



ACL LaTeCH 2016: August 11, Berlin

67299129912009_dectors-sected of 29612009_dectors-sected of 29612009_dectors-sected of 29612009_dectors-sect

INDEX

Introduction to Classic Egyptian

671893 and action accelerate all and acceleration accelerate a solar acceleration accele

- Encoding of hieroglyphic texts •
- System architecture
- Conclusions and future work



INDEX

Introduction to Classic Egyptian

671893 and action accelerate all and acceleration accelerate a solar acceleration accele

139610 Day decotton weeks at 1296 139 00 decotton weeks 1296 20 decotton weeks 26 20 and

- Encoding of hieroglyphic texts
- System architecture
- Conclusions and future work



or 2491200 march and a star and a star and a sector sector and a star and a star and a star a star and a star and a star and a star and a star a s

276 1 200 dector and a 1276 26 200 dector and 1276 26 200 dector and 1276 26 200 dector

CONTEXT

- Digital Heritage:
 - Use of computing and information technologies to study and preserve our cultural legacy
- Egyptology:
 Study of Ancient Egypt
- Goal: Text Information Retrieval (TIR) system for hieroglyphic texts

24 Starry dectron work of 129 Starry dectron work of 129 Starry dectron work of 129 Starry

HISTORY

- Longest-attested language:
 - 3300 BC 17c AD (productive) / today (Coptic Church)
 - Deep changes throughout its evolution
- Classic (a.k.a Middle) Egyptian:
 2100 BC 600 BC (spoken) / 5C AD (tradition)
 Stereotypical image of Ancient Egyptian

othe file Dath deethous weather the Alas Mark deethous weather the file Dath deethous weather the Alas Mark and

HISTORY

- Longest-attested language:
 - 3300 BC 17c AD (productive) / today (Coptic Church)
 - Deep changes throughout its evolution
- Classic (a.k.a Middle) Egyptian:
 2100 BC 600 BC (spoken) / 5C AD (tradition)
 Stereotypical image of Ancient Egyptian

Laste the and deel to a set to a share and a the share a share



orthe et and according the et and a section and the state of the et al a control according to the state of the all and a section according to the section according to the

CHARACTERISTICS

Afro-Asiatic language

- e.g., Arabic, Hebrew, Amharic, etc.
- Subfamily of its own

Consonantal

- Roots formed by consonants
- Only consonants are written

Pictographic

• Symbols portray elements of their world $\widehat{} \rightarrow$ a falcon

Philadon deethow we down a philadon deethow we do the philadon deethow we do the philadon

orteride and action weeks and arteride and account and a she and a she and a section weeks a she and a dechar weeks

139610 Day decotton weeks at 1296 139 00 decotton weeks 1296 20 decotton weeks 26 20 and

CHARACTERISTICS (cont)

No fixed alphabet
Evolved from 800 to 5,000 signs
New symbols/variants still appearing

671296-1009-dector-sected of 1.296-1009-dectors-sected of 1.296-1009-dectors-sected of 1.296-1009-dectors-secte

SIGN TYPES

1. Phonograms: represent (1-3) sounds (h) (s3) (hpr)

2. Logograms/ideograms: represent the depicted element

△: "eye" (irt)

3. Determinatives: not read; denote the semantic group of the word
: [MAN – HUMAN BEING]

orteriorandeerhoursectoreste

orthe et al and dealtow week of the Ala and dealtow week dorthe et al and market tow week ow week and a contract and the sector week ow week ow

ARRANGEMENT OF SIGNS

- **Continuous script:** no dividers to separate words or phrases **ADDED Continuous** (*iw 3pdw hr nht*) *"The birds are on the sycamore"*
- Arranged in non-linear groups
 - No formal rules but principles/heuristics (

1896 1 200 dector was to a she all a share to a she all a she

- Horror vacui
- Seeking harmony and aesthetics *"sycamore"* (*nht*) : $\square \square \bigcirc \bigcirc \bigcirc \rightarrow \square \square \bigcirc \bigcirc$

DIRECTION OF READING/WRITING

671893 and action accelerate all and acceleration accelerate a solar acceleration accele

ngelabon dector and ongelabon dechor and hone and hone and a contor and a ngelabor and a ngelabon dechor and a

Very flexibleNot fixed

Let's see some examples!

67293/2019_deethou-seeddollegialesta deethou-seeddollegialesta ei an deethou-seeddou-seeddollegialesta deethou-seeddo

DIRECTION OF READING/WRITING (cont)



orteristand dectron weather the state of a control dectron weather the state of the astrone weather a section was the control of the state of the st

67296/2009_dectore____dother addet and actions_____doctore. dother addet and actions_____dother addet and a dectore _____

DIRECTION OF READING/WRITING (cont)



one for a contraction and and the sector and a sector and the sector action and a sector action to a sector action and the sector action and the sector action action and the sector action action

67296-12000-dectore-sected of 292-1200-dectore-sected to 2000-dectore-sectore-sectore-sectore-sectore-sectore-sectore

DIRECTION OF READING/WRITING (cont)



MERCHAR dectors and a Merchan dechors and a Merchan and a Mark dectors and a Merchan

672963000 dector weater a fallor acchain weater and a fallor acchain acchain acchain acchain acchain acchain acchain acchain a fallor a fallo

DIRECTION OF READING/WRITING (cont)



The gland and a contract of the gland and a contract of the gland and a contract of the gland and a contract and

67293/2019_deethou-seeddollegialesta deethou-seeddollegialesta ei an deethou-seeddou-seeddollegialesta deethou-seeddo

DIRECTION OF READING/WRITING (cont)



other all a contract and a the contract and the contract and the contract and a contract a state of the contract and a contract a contract and a contract a contract

INDEX

Introduction to Classic Egyptian

orliggian deellow and a share a share and a share a sha

Achieven dechow wert do the state and decentor wert do the state to an decentor wert do the state and

- Encoding of hieroglyphic texts
- System architecture
- Conclusions and future work



671291200-dectore-endote91291200-dectore-endote91200-dectore-endote912000-dectore-endo

1896 Jacob dectow weeks and 1891 896 Jacob dechow weeks 1891 1891 Jacob and a the share weeks and the share and

GARDINER'S LIST

 Problem: how to represent hieroglyphs without drawing them

1 var. of this 'refuge'. 1) Fran. of ibs 'dance'.

Solution: encoding signs using regular characters

ortegilanon-dechoused dortegilanon-dechoused dortegilanon-dechoused dortegilanon-

GARDINER'S LIST (cont)

- Standard reference (723 signs + 20 var)
- Hieroglyphs encoded as characters:
 Sign code = category letter + seq. number:

= B2

- 26 categories according to drawings
- Numbered sequentially within category

	1	2	3	
A. "Man and his occupations":	Y	Ê	SA .	
B. "Woman and her occupations":	M	Æ	A.	•••
				1

276/200 dector dector and a har a follow dector and the har a follow have to a har a follow here have

othe eight and dechous and dothe eight and and down and do the eight and dechous and do the eight and all and a

GARDINER'S LIST (cont)

- Standard reference (723 signs + 20 var)
- Hieroglyphs encoded as characters:
 Sign code = category letter + seq. number:

= B2

- 26 categories according to drawings
- Numbered sequentially within category

A PARTY NOT		1	2	3	
1 10 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	A. "Man and his occupations":	R	A	A.	1
AL THE A	B. "Woman and her occupations":	Ň	A	A.	
					1

METLADA dector ecoldor Achara Achara Acentor ecoldor Acentor Acentor Acentor Acentor Acentor Acentor

MANUEL DE CODAGE

- Standard encoding for digitization
- Evolution of Gardiner's List:
 - Extra codes and rules for accurate representation of features (ASCII only)

Sign arrangement operators

Symbol	Operation	Example				
-	concatenation	Q3-X1-Z4-N1				
	subordination	X1:Z4:N1] %			
*	juxtaposition	Q3*X1:Z4				
()	grouping	Q3*(X1:Z4):N1				

steriland dechow-section for the state of a control dechow-section to state to a state to a state of the stat

she to a contract of the she and a contract of the she and a she and a she and a she and a contract of the she and a contr

MANUEL DE CODAGE (cont)

(b)

(c)

Damaged texts: special marks (*shades*) attached to sign codes



671893 and action accelerate all and acceleration accelerate a solar acceleration accele

6129212921200_deethou_eeddoll?921200_deethou_eeddoll?9212000_deethou_eeddoll?9212921292

INDEX

Introduction to Classic Egyptian

671893 and action accelerate all and acceleration accelerate a solar acceleration accele

- Encoding of hieroglyphic texts
- System architecture
- Conclusions and future work



67291291200 dector ender 201291291200 dector ender 201291200 dector ender 201291201200 dector

SYSTEM ARCHITECTURE



671296 and dectors and arte 96 and dectors and to 1296 and dectors and to 1296 and arter and to 1296 actors and

67291291200 dector ender 201291291200 dector ender 201291200 dector ender 201291201200 dector

SYSTEM ARCHITECTURE



671296 and dectors and arte 96 and dectors and to 1296 and dectors and to 1296 and a contain and to 1296 action and to

67291291200 dector welder 201291200 dector welder and to 2291291200 dector welder 201291291200 dector

TEXT NORMALIZATION

Regular text:

 Regular normalization process: standard tokenization, lowercasing, stopwords, etc.

Hieroglyphic text:

No delimiters !

 Initial approach: tokenized in sign groups (delimited by '-' in encoding):
 e.g. "boat" (dpt): ☐ ↓ D46:Q3*X1-P1
 → ☐ D46:Q3*X1

 $\rightarrow \checkmark$

LEP 1. DOAL doe HOW - sector and Plan and Acador - sector 1. EP 1. DOAL doe How - sector 1. EP 1. DOAL doe How -

P1

67291291200 dector ender 201291291200 dector ender 201291200 dector ender 201291201200 dector

SYSTEM ARCHITECTURE



671296 and dectors and arte 96 and dectors and to 1296 and dectors and to 1296 and arter and to 1296 actors and

67291291200-dectore-ender2929200-dectore-ender202912912000-dectore-ender2029200-enderender

SYSTEM ARCHITECTURE



FRONT-END INTERFACE

Family Y. Writings, games, music	File View Help
Sub-Family all	Search
Sub-Sub-Family	
✓ Show all filter signs containing	temple ramesses Approximate latin search
	t:Y1:3-D37:k-wn:n:z-mi:t:t-nTr*Hwt:k Approximate hiero search
RY R R	Group vertically Group Horizontal
	Cartouches Shades
	Results
	Luxor
	Ankhnesneferibre.odt
	King of Upper Egypt (Part 1).txt
	Ramses
	Temple of Philae.pdf
dmD. mDAt	Rosseta Stelae
	Osiris sarcofaque bie
V1	
Ĩ S S user Pal.	

671296 and dectors and arte 96 and dectors and arte 12 and dectors and arte 12 and a contain and a and a contain and a

67296/2009_dectore-sectore-sectore-dectore-secto

FRONT-END INTERFACE

Contraction of Para			CONT ONLY	A REPORT OF A				
Family	Y. Writi	ngs, games, m	usic		-	File View Help		
Sub-Family	all				•	╔╻╧┉╱╧╗╏┨╏		Search
Sub-Sub-Fami	ily				•	III 🗢 🛲 Š 🖳		Search
✓ Show all	filter signs containing Search			rch	temple ramesses		Approximate latin search	
~~~	k	<u> </u>	<u> </u>	٩		t:Y1:3-D37:k-wn:n:z-mi:t:t-nTr*	Hwt:k	Approximate hiero search
	ĴĴ	₽₹		Ĭ		Group vertically	Group Horizontal	
	, 		Δ	\mathbb{D}	=	Cartouches 🔻	Shades 🔻	
	盘	Ş	IE			Results		
Õ	Ϋ́	A	₩.			Luxor Ankhnesneferibre.odt	INF	PUT ARFA
	窶	<u>D</u>		Ļ	•	King of Upper Egypt (Part 1).tx Ramses	t	
						Temple of Philae.pdf		
	v	alues				Rosseta Stelae		
<u> </u>		amD, mDi	At			Templo de Filae		
V1						Osiris sarcotagus.hie		
	user Pal.	4			Mà € 1			
1 Col					P MA			
								151111

othe file Date deethous section of the Date and a start of the Date Aconton aconton acontain a section aconton acontain a section acontain a

67296/2009_dectore-sectore-sectore-dectore-secto

FRONT-END INTERFACE

And the second		and the second second		NAME AND ADDRESS OF TAXABLE PARTY.	1.11		
	N. 116 1					File View Help	
Family	Y. Writ	T. Writings, games, music					
Sub-Family	all	all				<u> </u>	Search
Sub-Sub-Famil	ly				•		Search
🗹 Show all	filter s	igns containing	I	▼ Search		temple ramesses	Approximate latin search
	×	<u> </u>	<u> </u>	<pre>d</pre>		t:Y1:3-D37:k-wn:n:z-mi:t:t-nTr*Hwt:k	Approximate hiero search
	ĴŢ	₽		Ť.			
¥₽			\bigtriangleup	\mathbb{P}		NPUT QUERY FOR	RIVIS
D	Å	¥	II			Results	
0	¥	А	ATTA			Luxor	
		\square	ĮЩĮ			Ankhnesneferibre.odt	
\square	窶	<u> </u>	Ð	Ļ	-	King of Upper Egypt (Part 1).txt	
					_	Ramses	
		,				Respects Stellag	
		dmD, mD	At			Templo de Filae	
						Osiris sarcofagus hie	
¥1							
	F						
2.2	user Pal.			Ր⇒ՎՓ			
		SC SCOTTER	Status - The later				
						经价格和价产量 带型管子系 和价产量人	
							SAL AL

othe file han deembourned to the file hand deembourned to the file hand deembourned to the file hand a contraction and contraction and the file hand a contrac

FRONT-END INTERFACE

							- BANK I III I A A		
Family	Y. Writings, games, music			-	File View Help				
Sub-Family	all			•					
Sub-Sub-Family				-	Search Search				
🗹 Show all	filter signs conta	ining	- Search		temple ramesses		Approximate latin search		
)			4		t:Y1:3-D37:k-wn:n:z-mi:t:t-nTr	*Hwt:k	Approximate hiero search		
	Î GÎ		- *		Group vertically	Group Horizontal			
				=	Cartouches 🔻	Shades -			
	₽ ₽		<i>U</i>		Results				
	і і ¥∕ д	Date: A	8		Luxor				
6		TUT			Ankhnesneferibre.odt				
	Ω	R .	Į,	_	King of Upper Egypt (Part 1).t	xt			
7 I I I I I I I I I I I I I I I I I I I	SP - 37	×	Ŷ	_	Ramses				
	_				Temple of Philae.pdf				
200	values	mD4+			Rosseta Stelae				
		, IIIDAt			Templo de Filae				
					Osiris sarcotagus.hie				
ΎΙ									
en 🗆 Z.a	ser Pal. 🗢		p→ Au						
					and the second second second				
							12 1 1 1		
SYM	BOL	PALFT	TF (JSF	SF	P STATISTICS OF				
			(

6712961200-dectors-sected of 2920200-dectors-sected 129200-dectors-sected of 2920200-dectors-sected

671291-1009-dectore-sectore-sectore-dectore-secto

FRONT-END INTERFACE

									INA COX - I DEB F D A A					
Family	Y. Writ	ings, games, m	usic		-		File View Help	ile View Help						
Sub-Family all									Search					
	(1), (1), (1), (1), (1), (1), (1), (1),						-							
Show all	TIITER S	igns containing]	Search			temple ramesses		Approximate latin search					
<u></u>	Å	_ ~ _	<u> </u>	4	^	h	Y1-2-D37 in an and the Tai		Approximate hiero search					
	ĴĴ	₽₹	₽	Ĭ.			Group vertically	Group Horizontal						
	, 		\square		=	L	Cartouches 🔻	Shades						
	盘	¥	E				Results							
ě	¥	Å	A∏∏A				Luxor SIG		ANGEWIEN					
	裏	<u> </u>	ा <u>स्</u> ता ष्ट्रि	ų.	•		King of Upper Egypt (Part 1).tx	t						
				· · · · · · · · · · · · · · · · · · ·			Ramses							
		values					Rosseta Stelae							
		dmD, mD,	At				Templo de Filae							
							Osiris sarcofagus.hie							
Y1														
	user Pal.	4		β÷	Ψħ									
									15111					

othe of a contract of the other and a start and the other and the start and the other and the other

FRONT-END INTERFACE

and the second second second	Contraction of the local sectors of the local secto					and the second second			
Family	Y. Writ	ings, games, п	nusic		•	File View	Help		
Sub-Family	all						1 JE 0 90		
Sub-Sub-Fam	ilv						▔▓░▋		Search
	····,								
Show all	filters	igns containing	3	▼ Sea	arch	temple rame	esses		Approximate latin search
~~~	×	<u></u>	ڪ	4	<u> </u>	t:Y1:3-D37:	k-wn:n:z-mi:t:t-nTi	r*Hwt:k	Approximate hiero search
பி	ĴŢ	ណ្ឌ	ណ្ឌ	۳L		Group	vertically	Group Horizontal	
	- 0	۸°۱	ĕ °	₽ ₽	=	Cart	ouches 🔻	Shades 🔻	
Ĩ.	<u></u>		\wedge	\square					
	STP	್		4		Results	SE/	аксн ілі	ODE SWIICH
2	Ŷ	Ŷ	TE						
õ	Ϋ́	д	Длтд			Luxor			
		\square	ΥЩΎ	, and the second		Ankhnesnef	eribre.odt		
\frown	垂	Ω	8	Ŧ	-	King of Upp	er Egypt (Part 1).t	xt	
	286	w/	•		·	Ramses			
						Temple of P	hilae.pdf		
		values dmD. mD	At			Rosseta Ste	lae		
		unit, int				Osiris sarcot	fagus bio		
N/1						USITIS SALCO	lagus.me		
T T	l.								
Za	user Pal.	4			β►₫Φ				
			Contraction of the last					and the first state of the second state of the	
									A 2 A A I I
	a star						基闭。(ED-11		TRITIN

61294312009_decttors-sected of 296312009_decttors-sected of 296312019_decttors-sected of 29632009_decttors-sected

FRONT-END INTERFACE

	and the second second						10.70	HIS NOX
Family	Y. Writ	ings, games, m	usic	-	1	File View Help		
Sub-Family	all			•		<u>₽ ▲ 4</u> €,090		
Sub-Sub-Fam	nily			•				Search
🗹 Show all	filter s	igns containing		Search]	temple ramesses		Approximate latin search
	×	_حك_	حکے	<u>م</u>		t:Y1:3-D37:k-wn:n:z-mi:t:t-nTr	*Hwt:k	Approximate hiero search
	ĴĴ	€Ĥ	₽			Group vertically	Group Horizontal	
			\wedge			Cartouches 🔻	Shades 👻	
	Â	¥				Results		
õ	Ϋ́	A	ATTA			Luxor		
		\square	ĬЩĬ			Ankhnesneferibre.odt		
\cap	鷾	Ê	B	Ļ –		King of Upper Egypt (Part 1).t>	t	
					-	Temple of Philae.pdf		
		values				Rosseta Stelae		
<u> </u>		dmD, mD,	At			Templo de Filae		
						Osiris sarcofagus.hie		
Y1								
Zà	user Pal.	\		₽►₫₽				
				OUT	F	PUT: RELI	EVANT D	OCUMENTS
	And the second second		A Designation of the second		10			

FRONT-END INTERFACE

					_	File View Heln	
Family	Y. Writ	ings, games, n	nusic		•		
Sub-Family	all				•	<u> </u>	Search
Sub-Sub-Fa	amily				•		
✓ Show al	l filter s	igns containing	9	Search		temple ramesses	Approximate latin search
	ł	<u> </u>	<u> </u>	¢	^	t:Y1:3-D37:k-wn:n:z-mi:t:t-nTr*Hwt:k	Approximate hiero search
	ĴĴ	₽Ĵ	₽			Group vertically Group Horizontal	
Ĩ			Δ	$\overline{\mathcal{D}}$	=	Cartouches Shades	
	Å.	ş				Results	
ě	¥	A	A∏TA			Luxor Ankhnesneferibre.odt	
	裏	<u> </u>		 	•	King of Upper Egypt (Part 1).txt	
		1			_	Temple of Philae.pdf	CONITENIT
<u> </u>		dmD, mD	At			Templo de Filae	
2/1						Osiris sarcofagus.hie	
Υ <u>γ</u>	🗌 user Pal.	4		îk ▲ 1)			
1 Col	L			P - 214			
	THE REAL						ABAAAA

othe of a contract of the other and a start and a start of the other and a start our and the other and a contract and a contract a start our and a contract of the other and a start our and a

67291/291/2009_dectore-ender2291/291/2009_dectore-ender22491/291/2009_dectore-ender246/291/291/291/2009_dectore-

FRONT-END INTERFACE

File View

Abydos temple of Ramesses II. p. 531-532.

NNNNE EXery long lacuna

VARANKENALENASIA VARANGKARASIAN VARANGKARASIANASINANGKARASIAN VARANGKARASIAN VARANGKARASIAN VARANGKARASIAN VARA



orteriorandeentoureedateriorandeentoureedateriorandeentoureedateriorandeentoureedatoureedateriorandeentoureed

67129iloan dector welder 129iloan dector welder 129iloan dector welder a commendation action action action weld

FRONT-END INTERFACE

File View Abydos temple of Ramesses II. p. 531-532. PW2244KEECOTVAC)CEMPDAECAAP NNNNE EXery long lacuna 2045 2045 A RESINDARY (RADIA & (EVANDARIA)

672896/2009_deethoursed dor 12896/2009_deethoursed dor 12896/2009_deethoursed dor 12896/2009_deethoussed

INDEX

Introduction to Classic Egyptian

orliggian deellow and a share a share and a share a sha

- Encoding of hieroglyphic texts
- System architecture
- Conclusions and future work



671291/2019_dectore____dot1291/291/2019_dectore____dectore____det1291/291/2019_dectore____dectore____dectore___

CONCLUSIONS

- <u>First</u> TIR system designed to manage Egyptian hieroglyphic texts
 - Language and writing system
 - Encoding
- Special care with front-end
 - Intuitive and easy to use
- Available at (free license, open source): https://github.com/estibalizifranjo/hieroglyphs

orthe et al and a contract of the other and a contract dorthe et al and marce to a contract of the other and the

27 War dechow weekow weekow weekow weekow weekow weekow dechow weekow weekow weekow weekow weekow weekow weekow

FUTURE WORK

Study other retrieval solutions:

- Retrieval models
- Conflation and matching mechanisms
- n-Gram based processing

Take advantage of similarities:

- Arabic, Hebrew, Japanese, Chinese, etc.
- Create evaluation corpora

Collaborators ?

Thank you very much! Questions?

https://github.com/estibalizifranjo/hieroglyphs

Third-Party Pictures:

- Slides no. 12-16 (cropped pic.): original by Olaf Tausch (*Oltau*) available at <u>https://commons.wikimedia.org/wiki/File:Edfu Tempel 42-2.jpg</u> under the Creative Commons Attribution 3.0 Unported license.
- Slide 18 (cropped pic.): example of a handwritten entry, then printed lithographically; obtained from: R.O. Faulkner (2006). "Concise Dictionary of Middle Egyptian". Griffith Institute.
- Slide 22 (cropped pic.): original by *Lord-of-the-Light*; available at <u>https://commons.wikimedia.org/wiki/File:Medinet Habou Cartouche.jpg</u> under the Creative Commons Attribution-Share Alike 3.0 Unported license.
- PPT template: obtained at

http://www.free-ppt-templates.com/2013/12/egyptian-ppt-template.html

ACL LaTeCH 2016: August 11, Berlin

ong the falle and a company and a share of the and a share and the share and a share and a share and a share and