

TIR over Egyptian Hieroglyphs

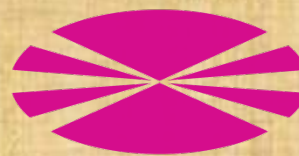
Estíbaliz Iglesias-Franjo, [Jesús Vilares](#)

Language in the Information Society Group
Universidade da Coruña

www.grupolys.org



TIR 2016: September 5, Porto



INDEX

- Introduction
- How to encode hieroglyphic texts
- Architecture of the system
- Conclusions and future work



INDEX

- **Introduction**
- How to encode hieroglyphic texts
- Architecture of the system
- Conclusions and future work



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**



CHARACTERISTICS

- **Classic (a.k.a Middle) Egyptian:**
 - 2100 BC – 600 BC (spoken) / 5C AD (tradition)
 - Stereotypical image of Egyptian
- **Afro-Asiatic language**
 - e.g., Arabic, Hebrew, Amharic, etc.
 - Subfamily of its own
- **Consonantal**
 - Roots formed by consonants
 - Only consonants are written



CHARACTERISTICS

- **Classic (a.k.a Middle) Egyptian:**
 - 2100 BC – 600 BC (spoken) / 5C AD (tradition)
 - **Stereotypical image of Egyptian**
- **Afro-Asiatic language**
 - e.g., Arabic, Hebrew, Amharic, etc.
 - Subfamily of its own
- **Consonantal**
 - Roots formed by consonants
 - Only consonants are written



CHARACTERISTICS (cont)

- **Pictographic**

- Symbols portray elements of their world



→ a falcon

- **No fixed alphabet**

- Evolved from 800 to 5,000 signs
- New symbols/variants still appearing



SIGN TYPES

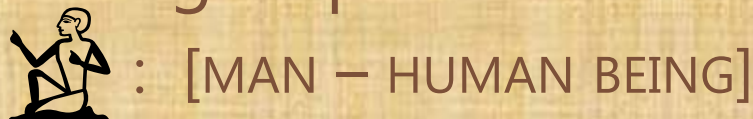
1. **Phonograms:** represent (1-3) sounds



2. **Logograms/ideograms:** represent the depicted element



3. **Determinatives:** not read; denote the semantic group of the word



ARRANGEMENT OF SIGNS

- **Continuous script:** no dividers to separate words or phrases

(iw 3pdw hr nht)

"The birds are on the sycamore"

- **Arranged in non-linear groups**
 - No formal rules but principles/heuristics
 - *Horror vacui*
 - Seeking harmony and aesthetics

"sycamore" (nht) : →



DIRECTION OF READING/WRITING

- Very flexible: four possible ways



INDEX

- Introduction
- **How to encode hieroglyphic texts**
- Architecture of the system
- Conclusions and future work



PROBLEM




"I'm glad they set that 140-character limit..."









- Solution: **encoding signs using characters**

GARDINER'S LIST

- 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":				...
B. "Woman and her occupations":				...
...









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":				...
B. "Woman and her occupations":				...
...



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 



MANUEL DE CODAGE (cont)

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



(a) <-N5-F12*C10-N36-M17*(Y5:N35)->

(c) <-N5-(F12#13)*C10-N36#13-M17*(Y5:N35)->

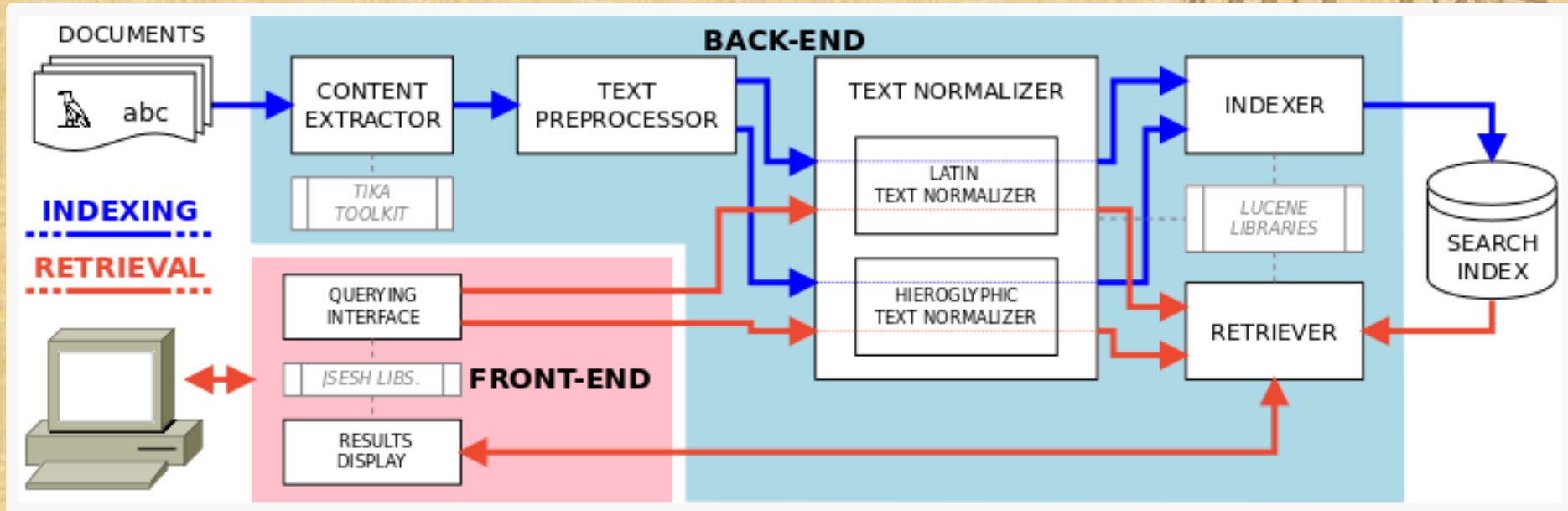


INDEX

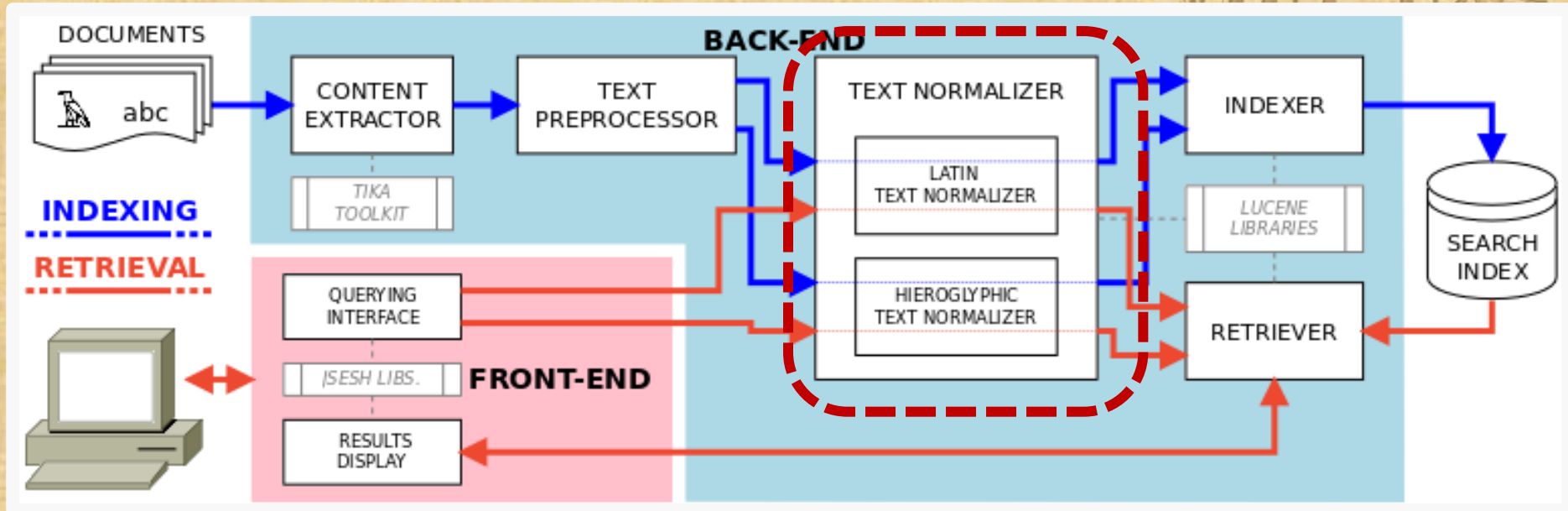
- Introduction
- How to encode hieroglyphic texts
- **Architecture of the system**
- Conclusions and future work



SYSTEM ARCHITECTURE



SYSTEM ARCHITECTURE



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):

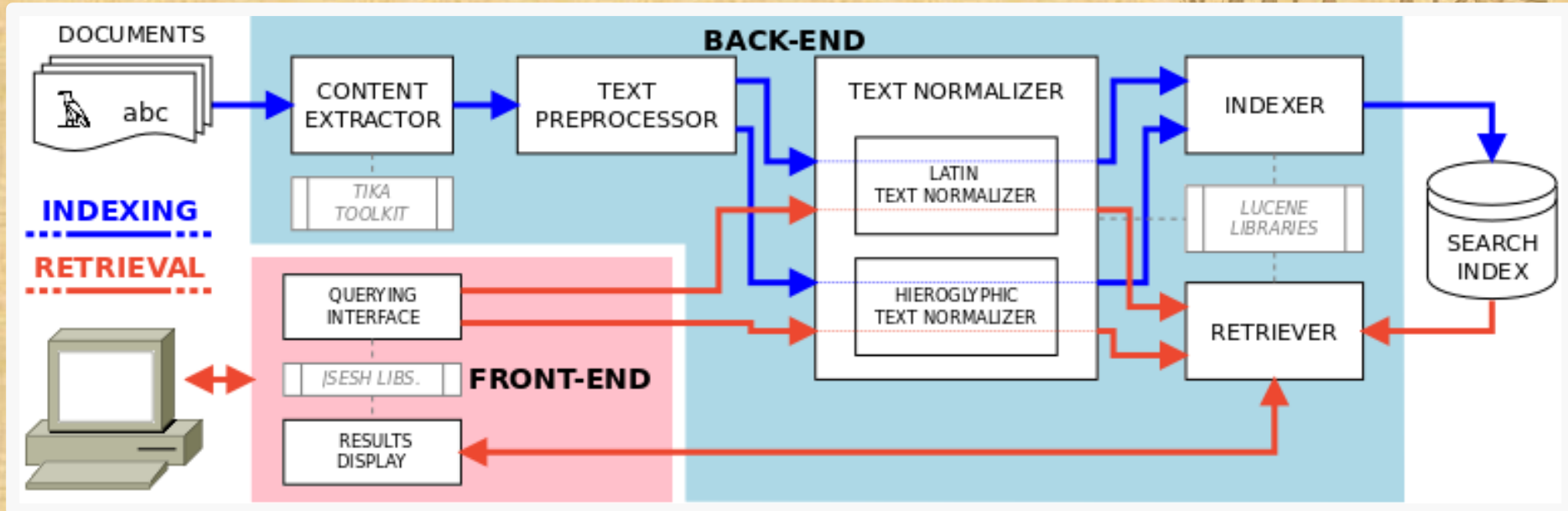
e.g. "boat" (*dpt*):  D46:Q3*X1-P1

→  D46:Q3*X1

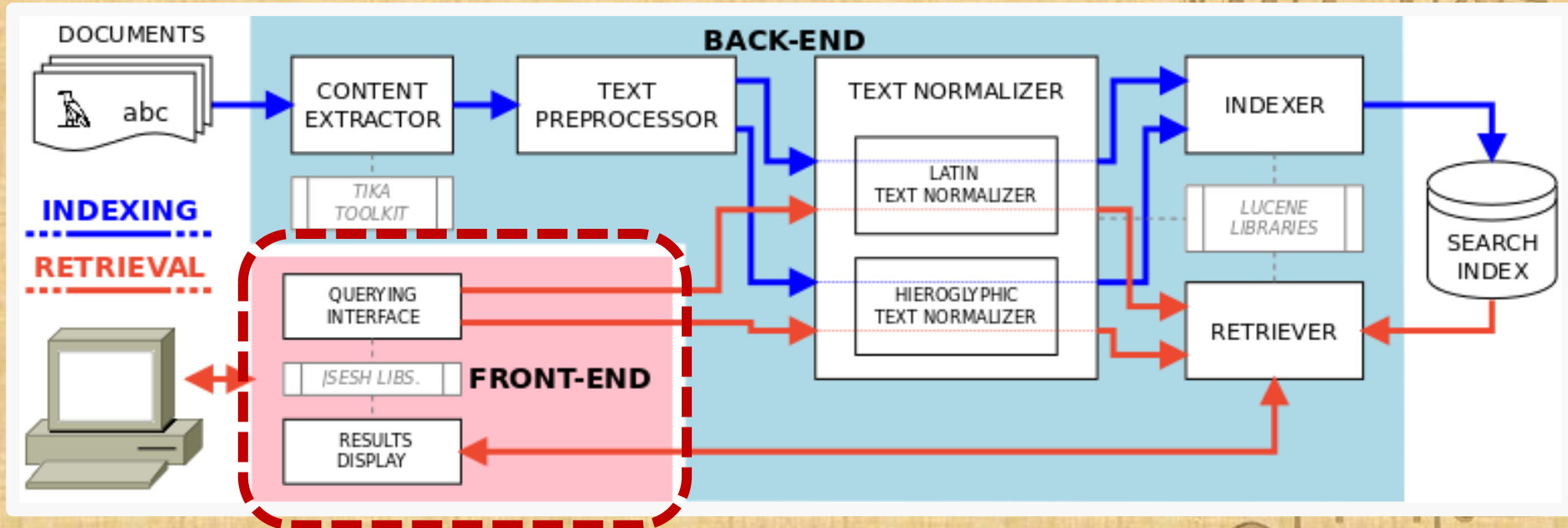
→  P1



SYSTEM ARCHITECTURE



SYSTEM ARCHITECTURE



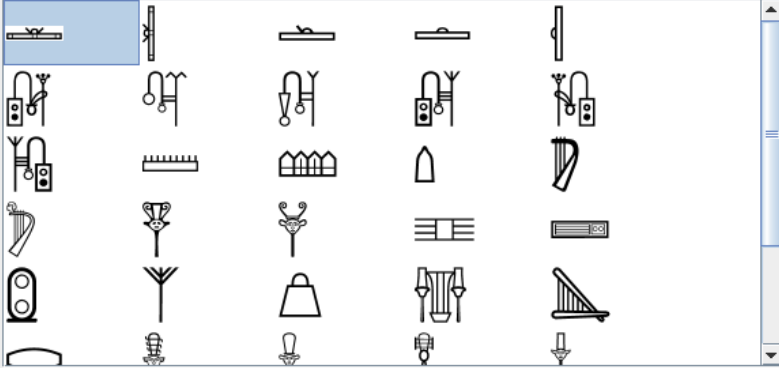
FRONT-END INTERFACE

The image shows a software interface for working with Egyptian hieroglyphs. It is divided into several sections:

- Left Panel (Sign Palette):** Contains a grid of hieroglyphs. At the top, there are dropdown menus for "Family" (set to "Y. Writings, games, music"), "Sub-Family" (set to "all"), and "Sub-Sub-Family". Below these is a "Show all" checkbox and a "filter signs containing" dropdown menu with a search input field. A large grid of hieroglyphs is displayed, with the first one selected. Below the grid is a preview window showing the selected hieroglyph (Y1) and its "values" (dmD, mDA). At the bottom left, there is a "user Pal." checkbox and a navigation arrow. At the bottom right, there are icons for undo and redo.
- Top Panel (Menu and Search):** Features a menu bar with "File", "View", and "Help". Below the menu is a search bar with the text "temple rameses" and a "Search" button. To the right of the search bar are two checkboxes: "Approximate latin search" and "Approximate hiero search".
- Middle Panel (Tools):** Contains several buttons and dropdown menus: "Group vertically", "Group Horizontal", "Cartouches", and "Shades".
- Right Panel (Results):** Titled "Results", it displays a list of search results: Luxor, Ankhnesneferibre.odt, King of Upper Egypt (Part 1).txt, Ramses, Temple of Philae.pdf, Rosseta Stelae, Templo de Filae, and Osiris sarcophagus.hie.


FRONT-END INTERFACE


Family: Y. Writings, games, music
Sub-Family: all
Sub-Sub-Family:
 Show all filter signs containing Search




values
dmD, mDat

Y1

user Pal. 



File View Help



Search

Approximate latin search
 Approximate hiero search

temple rameses

t:Y1:3-D37:k-wn:n:z-mi:t:t-nTr*Hwt:k

Group vertically Group Horizontal

Cartouches Shades

Results

- Luxor
- Ankhesneferibre.odt
- King of Upper Egypt (Part 1).txt
- Rameses
- Temple of Philae.pdf
- Rosseta Stelae
- Templo de Filae
- Osiris sarcophagus.hie

INPUT AREA

FRONT-END INTERFACE

The image displays a software interface for working with Egyptian hieroglyphs. On the left is a sign palette with a grid of hieroglyphs. The top of the palette shows settings for 'Family' (Y. Writings, games, music), 'Sub-Family' (all), and 'Sub-Sub-Family'. A 'Show all' checkbox is checked, and a filter is set to 'filter signs containing'. A search box is present. Below the grid, a selected hieroglyph is shown in a larger view, labeled 'Y1', with a 'values' field containing 'dmD, mDat'. On the right is a search window with a menu bar (File, View, Help) and a toolbar. The search window has a search bar containing 'temple ramesses' and a text input field with 't:Y1:3-D37:k-wn:n:z-mi:t:t-nTr*Hwt:k'. There are checkboxes for 'Approximate latin search' and 'Approximate hiero search'. Below the search window, a 'Results' list shows several items: Luxor, Ankhnesneferibre.odt, King of Upper Egypt (Part 1).txt, Ramses, Temple of Philae.pdf, Rosseta Stelae, Templo de Filae, and Osiris sarcofagus.hie. A large red dashed box highlights the search window and the text 'INPUT QUERY FORMS' is overlaid in red.

Family: Y. Writings, games, music
Sub-Family: all
Sub-Sub-Family:
 Show all
filter signs containing Search
Y1
values: dmD, mDat
user Pal.

File View Help
Search
temple ramesses Approximate latin search
t:Y1:3-D37:k-wn:n:z-mi:t:t-nTr*Hwt:k Approximate hiero search
Results
Luxor
Ankhnesneferibre.odt
King of Upper Egypt (Part 1).txt
Ramses
Temple of Philae.pdf
Rosseta Stelae
Templo de Filae
Osiris sarcofagus.hie

INPUT QUERY FORMS

FRONT-END INTERFACE

The image displays the front-end interface of the JSESH (Jython-based Egyptian Signatures) software. The interface is divided into several sections:

- Symbol Palette (Left):** A grid of Egyptian hieroglyphs. A red dashed box highlights this area. Below the grid, a selected hieroglyph is shown with its values: `dmD, mDAat`. The label `Y1` is visible below the values.
- Search Panel (Right):** A search interface with a menu bar (File, View, Help), a search bar, and a list of results. A red arrow points from the search bar to the symbol palette.
- Search Results (Right):** A list of search results including: Luxor, Ankhnesneferibre.odt, King of Upper Egypt (Part 1).txt, Ramses, Temple of Philae.pdf, Rosseta Stelae, Templo de Filae, and Osiris sarcophagus.hie.

SYMBOL PALETTE (JSESH)

FRONT-END INTERFACE

The screenshot displays a software interface for working with Egyptian hieroglyphs. On the left, a sign palette is organized into a grid. The top section includes filters for 'Family' (Y. Writings, games, music), 'Sub-Family' (all), and 'Sub-Sub-Family'. A 'Show all' checkbox is checked, and a search box is labeled 'filter signs containing'. Below the grid, a selected sign is shown with its 'values' (dmD, mDat) and the label 'Y1'. At the bottom left, there is a 'user Pal.' checkbox and navigation arrows.

The right side of the interface features a menu bar with 'File', 'View', and 'Help'. Below it is a search bar containing 'temple rameses' and a 'Search' button. Two search options are listed: 'Approximate latin search' and 'Approximate hiero search', both with unchecked checkboxes. A red dashed box highlights a set of arrangement controls: 'Group vertically', 'Group Horizontal', 'Cartouches', and 'Shades'. Below these is a 'Results' list containing the following items:

- Luxor
- Ankhnesneferibre.odt
- King of Upper Egypt (Part 1).txt
- Ramses
- Temple of Philae.pdf
- Rosseta Stelae
- Templo de Filae
- Osiris sarcofagus.hie

SIGN ARRANGEMENT

FRONT-END INTERFACE

The screenshot displays a software interface for working with Egyptian hieroglyphs. On the left, a panel shows a grid of hieroglyphs, with the 'Y1' symbol selected. Below the grid, a 'values' box lists 'dmD, mDat'. The top of the interface includes dropdown menus for 'Family' (Y. Writings, games, music), 'Sub-Family' (all), and 'Sub-Sub-Family'. A 'Show all' checkbox and a search box are also present.

On the right, a search window is open, featuring a menu bar (File, View, Help) and a toolbar with icons for search and editing. The search input field contains 'temple ramesses', and the results field shows a list of search results. A red dashed box highlights the search mode options: 'Approximate latin search' and 'Approximate hiero search', both of which are currently unchecked. The text 'SEARCH MODE SWITCH' is overlaid in large red letters on the right side of the interface.

File View Help

Search

Approximate latin search

Approximate hiero search

SEARCH MODE SWITCH

Results

- Luxor
- Ankhesneferibre.odt
- King of Upper Egypt (Part 1).txt
- Ramses
- Temple of Philae.pdf
- Rosseta Stelae
- Templo de Filae
- Osiris sarcophagus.hie

FRONT-END INTERFACE

Family: Y. Writings, games, music
Sub-Family: all
Sub-Sub-Family:
 Show all filter signs containing Search

File View Help

Search

temple ramesses

t:Y1:3-D37:k-wn:n:z-mi:t:t-nTr*Hwt:k

Group vertically Group Horizontal

Cartouches Shades

Results

- Luxor
- Ankhnesneferibre.odt
- King of Upper Egypt (Part 1).txt
- Ramses
- Temple of Philae.pdf
- Rosseta Stelae
- Templo de Filae
- Osiris sarcofagus.hie

OUTPUT: RELEVANT DOCUMENTS

FRONT-END INTERFACE

The image shows a software interface for accessing content, divided into two main panels.

Left Panel (Sign Selection):

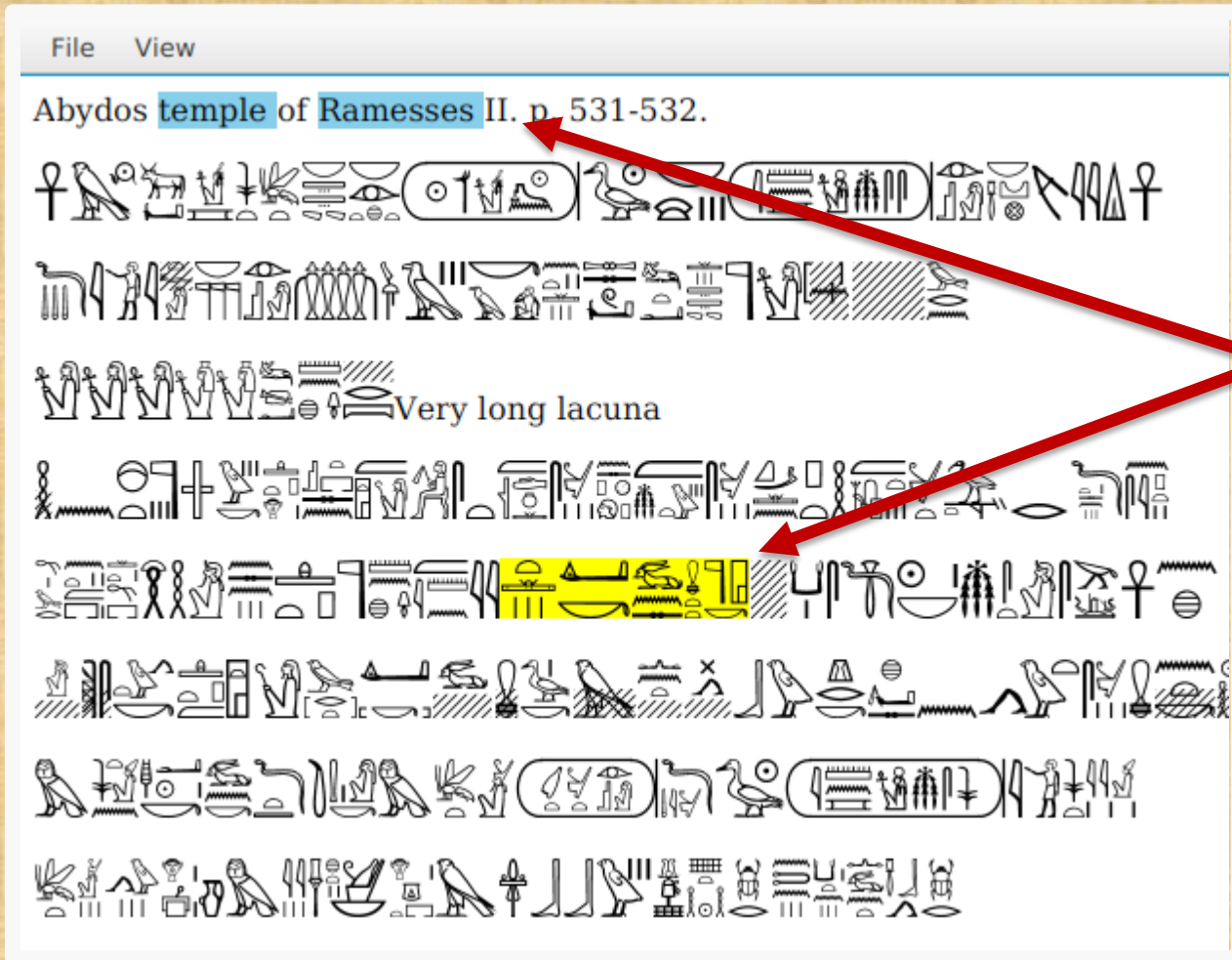
- Family:** Y. Writings, games, music
- Sub-Family:** all
- Sub-Sub-Family:** (empty)
- Show all:** filter signs containing [dropdown] Search [input]
- Sign Grid:** A grid of various Egyptian hieroglyphs. The first sign in the first row is highlighted with a blue background.
- Selected Sign:** A larger view of the selected sign (Y1) is shown below the grid.
- Values:** A text box containing the values "dmD, mDat".
- Y1:** A label for the selected sign.
- Buttons:** "user Pal." (checkbox), a left arrow, and a right arrow with a sign icon.

Right Panel (Search and Results):

- Menu:** File View Help
- Toolbar:** Icons for search, group, and other actions.
- Search:** Search [input]
- Search Text:** temple rameses
- Advanced Search:** t:Y1:3-D37:k-wn:n:z-mi:t:t-nTr*Hwt:k
- Buttons:** Group vertically, Group Horizontal, Cartouches [dropdown], Shades [dropdown]
- Results:** A list of search results:
 - Luxor** (highlighted with a yellow background and a hand icon pointing to it)
 - Ankhesneferibre.odt
 - King of Upper Egypt (Part 1).txt
 - Ramses
 - Temple of Philae.pdf
 - Rosseta Stelae
 - Templo de Filae
 - Osiris sarcofagus.hie

ACCESING CONTENT...

FRONT-END INTERFACE



MATCHINGS

INDEX

- Introduction
- How to encode hieroglyphic texts
- Architecture of the system
- **Conclusions and future work**



CONCLUSIONS

- **First 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>



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 an evaluation corpora**



Thank you very much! Questions?

<https://github.com/estibalizifranjo/hieroglyphs>

TIR 2016: September 5, Porto



DIRECTION OF READING/WRITING

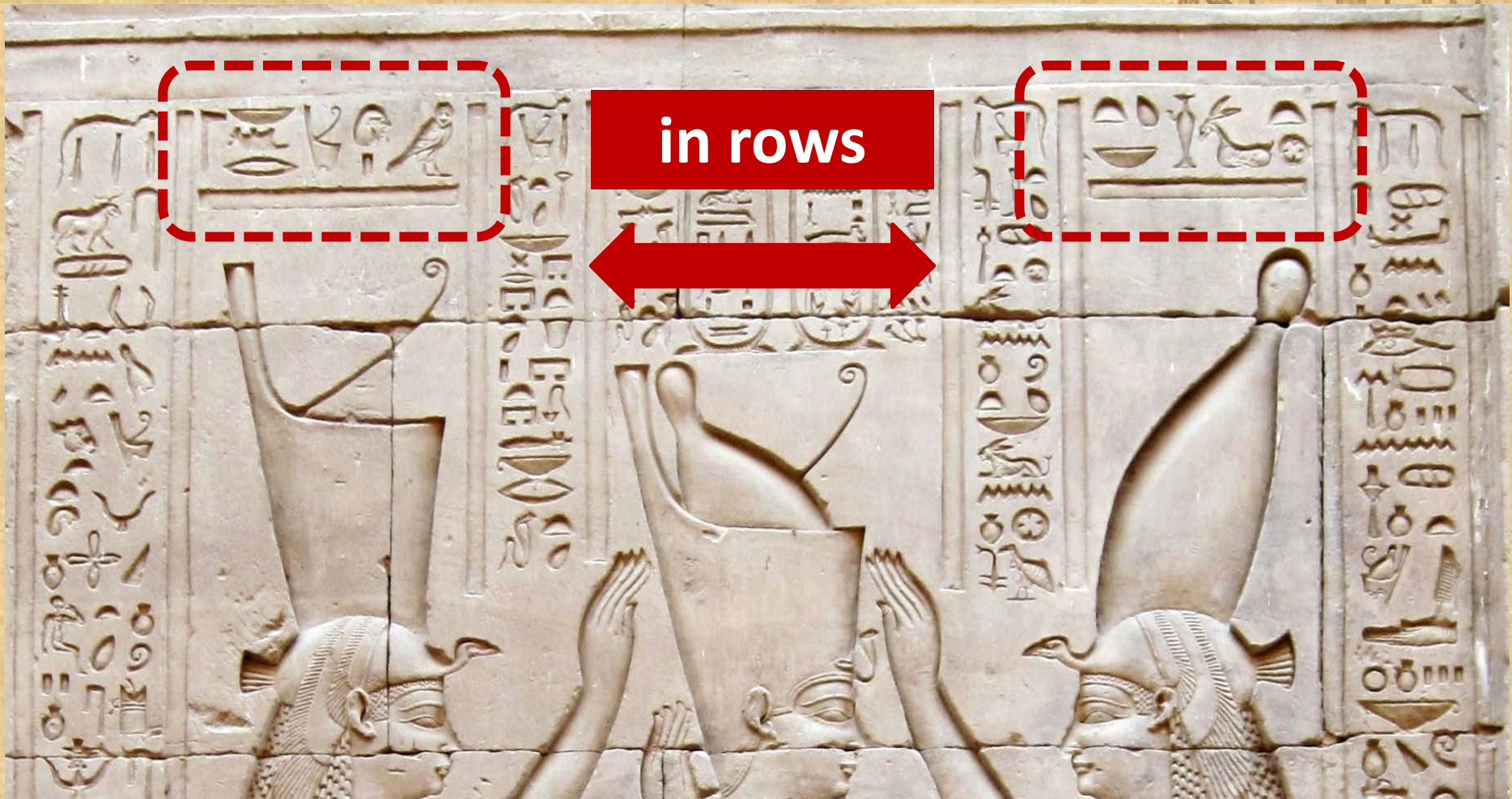
- Very flexible
- Not fixed
- Let's see some examples!



DIRECTION OF READING/WRITING (cont)



DIRECTION OF READING/WRITING (cont)



DIRECTION OF READING/WRITING (cont)



in columns

DIRECTION OF READING/WRITING (cont)

from-left-to-right



DIRECTION OF READING/WRITING (cont)



from-right-to-left

RELATED WORK

- Closely linked to the development of **classic-style text editors**:
 - GLYPH (1986): MS DOS, Windows, Mac
 - Laid the foundations of future editors
 - JSesh (2014): open source, Java
 - Currently the most popular



RELATED WORK (cont)

- **NLP/Text Mining:**
 - (Very) Initial stages: lack of corpora
 - Automatic transliteration (Barthélemy and Rosmorduc, 2011)
 - Language modelling (Nederhof and Rahman, 2015a)
 - Text categorization (Gohy et al., 2013)

