

Linked open data and machine learning applications for cuneiform studies



Timo Homburg

timo.homburg@hs-mainz.de, Hochschule Mainz, University Of Applied Sciences
funded by DFG grant 424957759: "Die digitale Edition der Keilschrifttexte aus Haft Tappeh"

References

[1] Timo Homburg. Paleocodage—enhancing machine-readable cuneiform descriptions using a machine-readable paleographic encoding. *Digital Scholarship in the Humanities*, 36(Supplement_2):ii127–ii154, 2021.

[2] Timo Homburg. Paleordia: Semantically describing (cuneiform) paleography using paleographic linked open data. 2024.

[3] Hubert Mara and Timo Homburg. MaiCuBeDa Hilprecht - Mainz Cuneiform Benchmark Dataset for the Hilprecht Collection, 2023.

[4] Ernst Stötzner, Timo Homburg, Jan Philipp Bulenkamp, and Hubert Mara. R-cnn based polygonalwedge detection learned from annotated 3d renderings and mapped photographs of open data cuneiform tablets. 2023.

[5] Ernst Stötzner, Timo Homburg, and Hubert Mara. Cnn based cuneiform sign detection learned from annotated 3d renderings and mapped photographs with illumination augmentation. In *Proceedings of the IEEE/CVF*, pages 1680–1688, 2023.

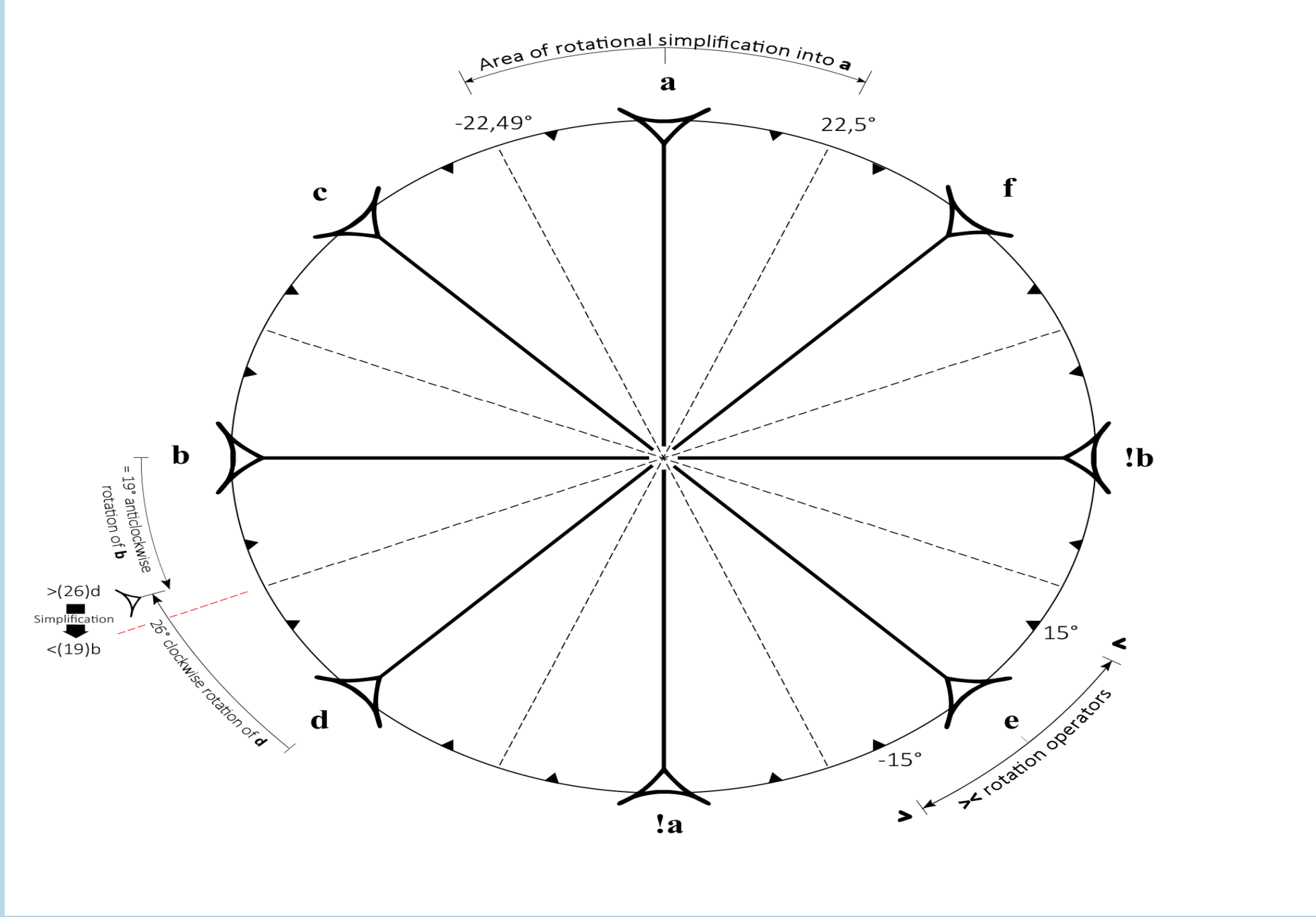
Motivation: Cuneiform Paleography

Idea: Representing cuneiform paleography for sign recognition and further machine learning tasks

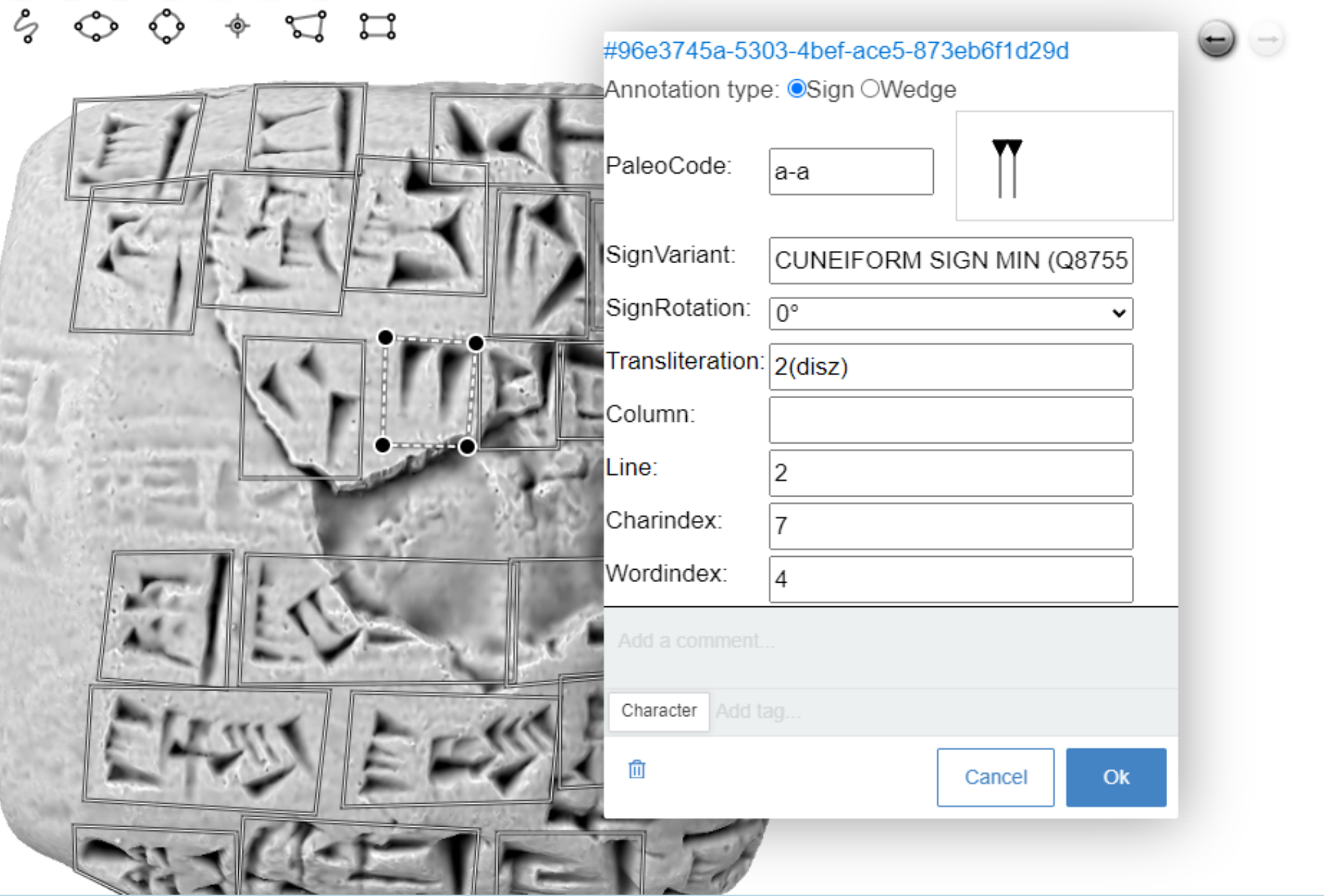
Time Period Designation Point in time	Late Uruk ca. 3100	Djemdet Nasr ca. 3000	Early Dynastic III ca. 2400	Ur III ca. 2000	Old Assyrian ca. 1900	Old Babylonian ca. 1700	Middle Assyrian ca. 1200	Neo- Babylonian ca. 600	meaning
Glyph (U+12295)									SAG "head"
Gottstein	N/A	N/A	a3b2c1d1	a3b2c1d1		a3b4	a3b3c1	a3b4	
Glyph (U+120FB)									NINDA "ration"
PaleoCode Gottstein	N/A N/A	N/A N/A	:sa~a~c~;<F a2c1d1	a~:sa~sa~::~f a3d1	sa~sa;w~sa a3c1	sa~:sa;w~sa a3c1	sa~sa;w~sa a3c1	f~;>c~:sb~a a1c2d1	
Glyph (U+12165)									GU ₇ "to eat"
Gottstein		N/A	a5b2c6d2	a5b2c4d2		a6b7c2d1	a5b5c1	a5b4c2	

From encoding cuneiform and annotating cuneiform to creating cuneiform benchmark datasets

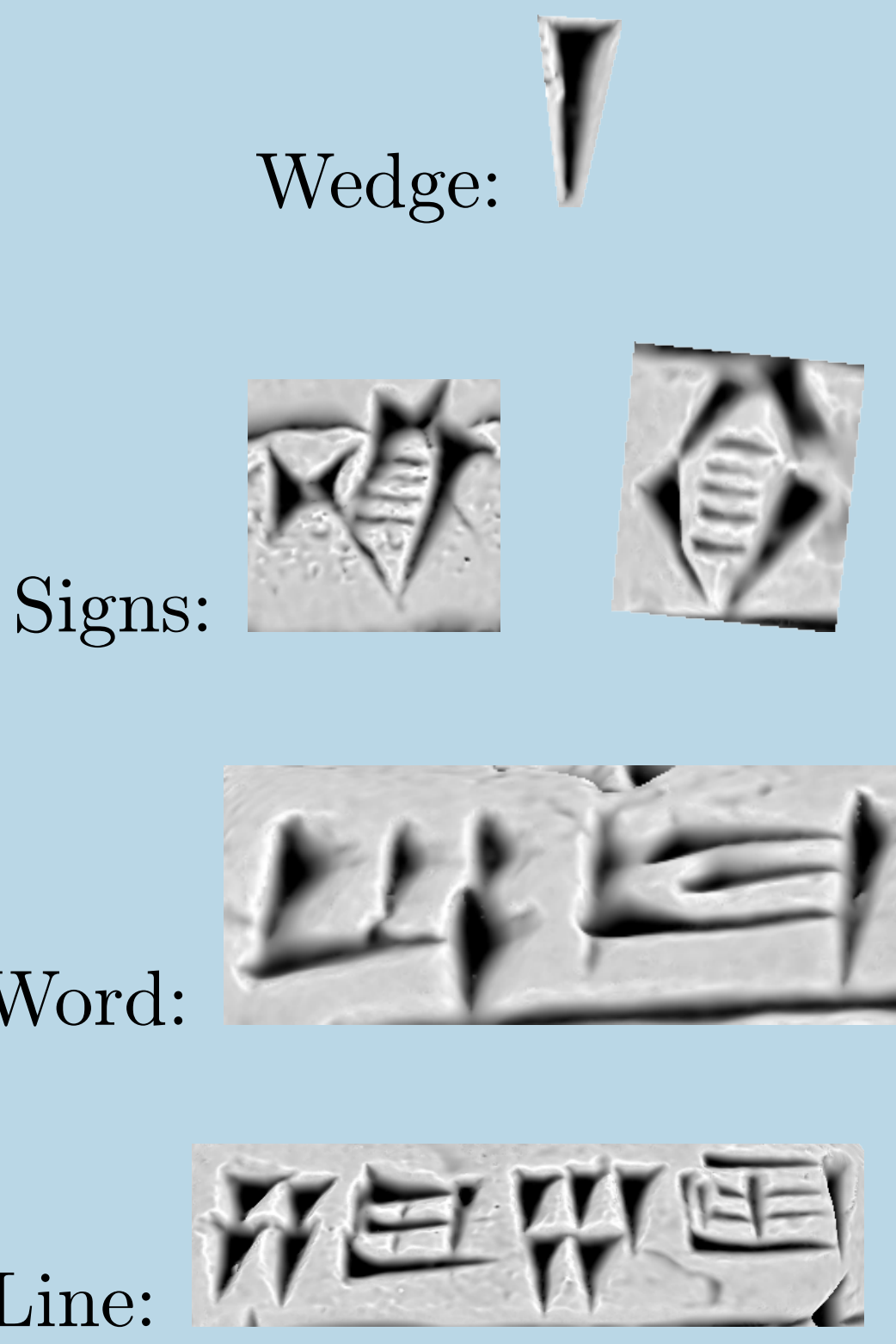
Encoding Cuneiform Signs with the PaleoCodage encoding [1]



Sign Annotation using the Cuneiform Annotator



Benchmark Dataset MaiCuBeDa [3]



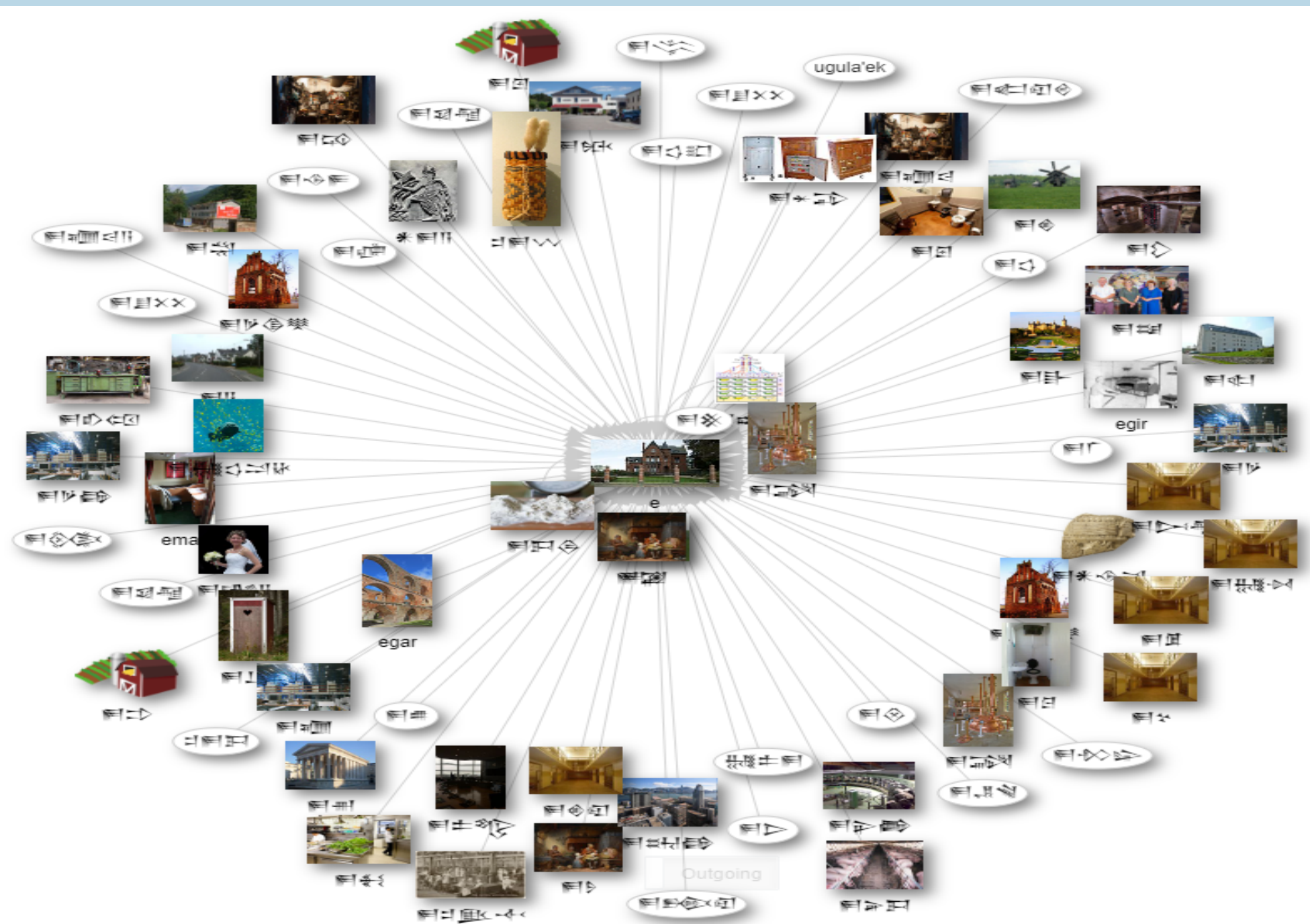
Cuneiform Linked Open Data Cloud

Cuneiform LOD Cloud

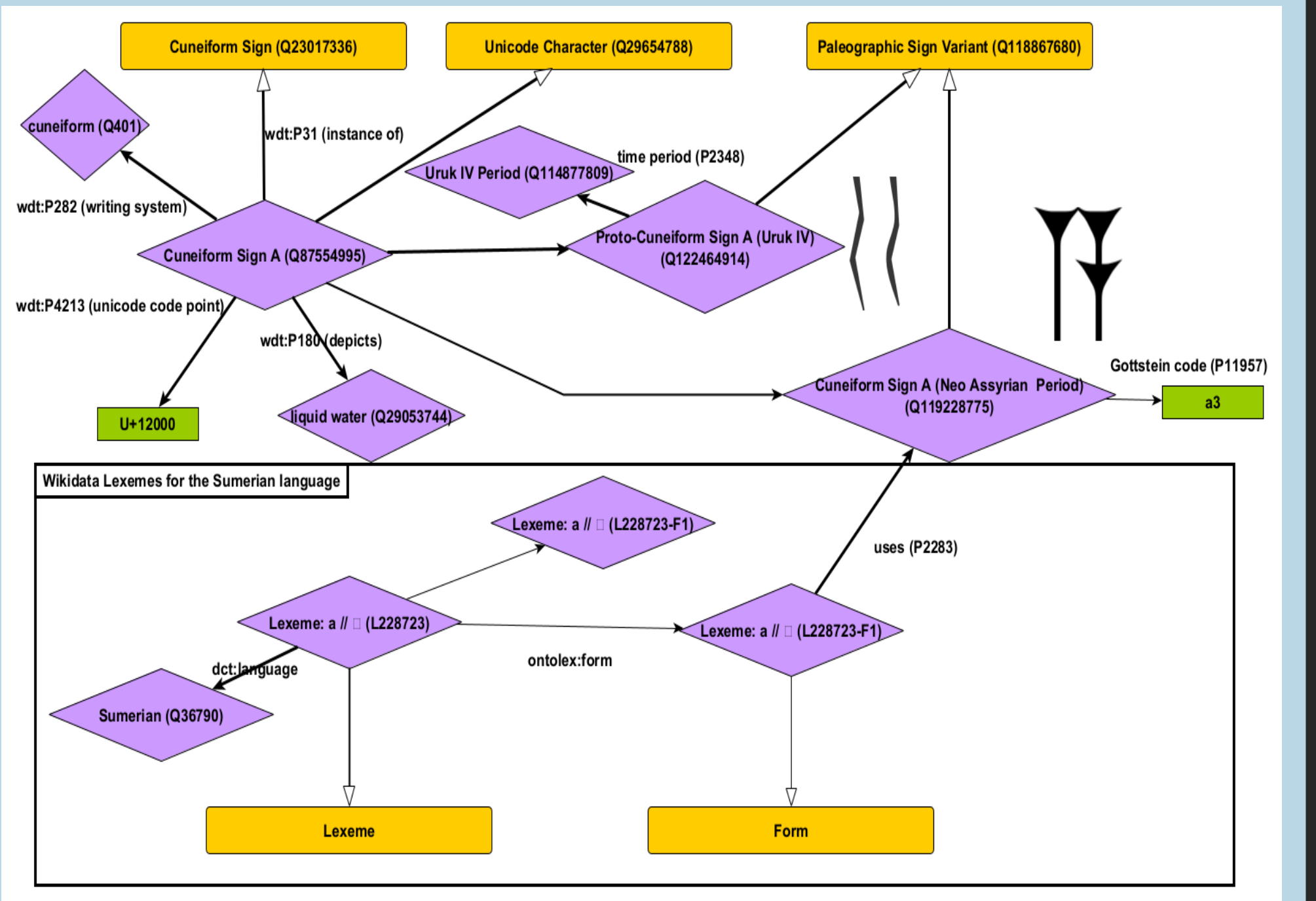
Capturing information about

- Cuneiform Signs and Paleography
- Linked Open Data Dictionaries
- Annotation data and metadata
- Artifact and media metadata
- Part Of Speech Tags
- Interconnections between Paleography and Dictionary Data

LOD Sumerian Dictionary [2]



Paleography Knowledge Graph



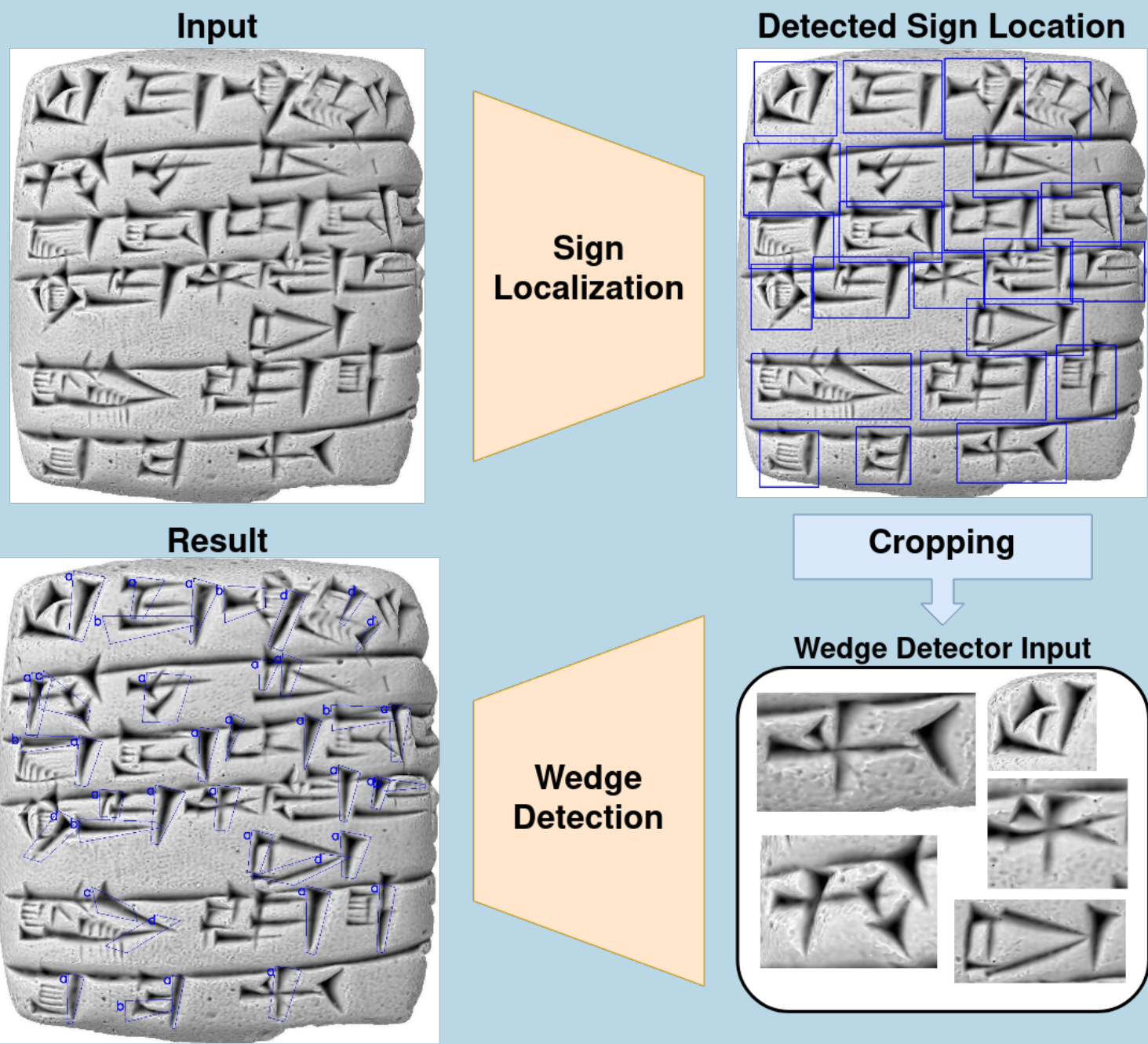
Machine Learning Classifications for sign and wedge recognition

Machine Learning Tasks

Sign recognition:

- On Different media (3D, renderings, photos)
- By detecting single cuneiform wedges
- Using different machine learning methods (e.g. CNN, RandomForest)
- Using reproducible datasets like MaiCuBeDa

Cuneiform Sign Detection [5]



Cuneiform Wedge Detection [4]

