ADVANCES IN KNOWLEDGE ORGANIZATION
Proceedings of International ISKO Conference - ISSN: 0938-5495 - 2026 - São Paulo

Real Characters as Geometric Semantic Notation in Knowledge Organization

John Kausch

Accenture, United States

Abstract

This paper describes two ontologies, one a vocabulary called Horapollo, and the other the Real Character Language, which reuses the Horapollo vocabulary for a fuzzy logic indexing system. The specific affordances of Real Character Language, and its relationship to the concept of semantic primitives, and its development from the CLLD Concepticon are described. A method for prototype theoretic indexing in a vector space produced by machine learning using cosine similarity is discussed. The specific design of the system, involving a fuzzy Boolean algebra, and a metric that allows the user to control serendipity, is outlined. Reflections for future work include integrating knowledge graphs with vector databases, and possible studies with human participants. A URL to the system is provided.

1.0 Introduction

This paper describes an act of research creation. The result of this research creation is published at the URL https://realcharacterlanguage.world. This project poses the question: how can we create pictograms which are highly translatable into many different languages, and use them to index documents? This paper provides a theoretical background which situates my work of digital art within a particular theoretical context. This context, in turn, informs my work as an intervention within theoretical domains essential to library and information science. It asks the question: what happens to knowledge organization when we embrace ambiguity in our designs, rather than attempting to standardize it out of existence?

One of the methods used for my work is research creation. Research creation is a mode that uses creative or artistic practice to intervene in the academy, and to produce art objects whose situated existence challenges institutional or theoretical norms and values (Loveless 2015, 2019), often transcending strict disciplinary boundaries (McCormack 2008). Research creation art objects both embody research questions and inspire new ones (Manning 2016, Proposition 3), using the tools of creative process to reflect on findings and disseminate results for knowledge translation, and they also transform how those results are represented and understood. Most importantly, research creation takes artistic methods as a generative practice for producing new knowledges which often resist simple disciplinary classifications: research creation is itself a mode of producing knowledge which Manning (2016) calls philosophy as a practice. In this sense, research creation engages in a transformation of the epistemic media of research itself, adding to knowledge in a field or fields through creative practice.

Here, the art object produced, and the practice interrogated, is a knowledge organization system (KOS) which consists of two ontologies. Knowledge organization systems are cognitive tools that index documents using pre-defined terms, for the purpose of helping people locate, access and retrieve documents (Hjørland 2008; Tennis 2008). Traditionally affiliated with library and information science, today the field of knowledge organization also incorporates elements of computer science, philosophy and linguistics (Gnoli 2025). An archetypal knowledge organization system is a library classification system (such as the Dewey Decimal Classification), but they can also be glossaries, taxonomies, thesauri and ontologies, used in computer systems for information retrieval, or in special sciences (such as the Linnaean Taxonomy in biology) (Mazzocchi 2018; Zeng 2008).

The object of creating a KOS which is also an art object is to invite questions into the nature of indexing in information access, and, in this case, how cross-linguistic needs and concerns intersect with information science. The system that is produced here is a fully functioning KOS. It is designed to index documents in multiple languages using the same vocabulary. It consists of two sub-ontologies, one called Horapollo, and the other called Real Character Language.

The term Real Character Language is inspired by Wilkins (1668, 1968). A Real Character Language is a language of images, or pictograms, and the term is based on an Enlightenment-era distinction between the hieroglyphic tradition of the Renaissance and Francis Bacon's (1605, [190?]) notion of "Real Characters." As such, my vocabulary is a wordlist of translatable concepts. Since these many invented language schemes of the 17th century consist of wordlists, and in some cases, taxonomies, following Gnoli we can interpret such systems as simultaneously conlangs (or constructed languages) and as knowledge organization systems (Gnoli 2018).

2.0 Features

2.1 Semantic Primitives

This KOS intervenes in the debate between phenomenon-based and domain-specific classification systems by embodying a different paradigm. The purpose of its classes is not to exhaustively describe reality, nor is it a top-level ontology in any way. Rather, the classes are meant to be understood as representing terms which are highly translatable. The purpose of this system is thus to organize information across different vector databases and provide a standard whose classes are more likely to be reusable in a large number of domains.

Each class in the KOS is represented by a single character in the alphabet, and given a dedicated visual representation. The purpose of this visual representation is to be unambiguous, and follows the general outline laid out by Kausch (2024). The use of icons does not fully eliminate ambiguity. However, it does seek to aid interpretability, by providing unambiguous representations of common objects that are highly translatable, in order to facilitate a shorthand which can be reused in several languages.

The concepts and their icons form the ontology Horapollo, which is implemented as a simple font, installable on any machine. The ontology is implemented in OWL, and links each image in the font to IPA phonemes and Unicode characters that are both visually and phonetically similar to the concepts in Horapollo. The vocabulary of Horapollo is reused to define the Real Character Language ontology, which arranges these concepts in hierarchical relations to each other, and also in relationships of opposition, through the use of fuzzy Boolean operators (Zadeh 1988), realized in OWL as owl:intersection, :union and :negation.

2.2 Concepticons

A concepticon is a word list of concepts which are highly translatable crosslinguistically (Poornima and Good 2010). In this sense it can be thought of as a meta-wordlist, as it presents a set of concepts with a statistically prominent distribution of potentially translatable words in all the world's languages. The first concepticons emerged through the linguistic research of Morris Swadesh (1955), whose glottochronology sought to statistically assess cross-linguistic universals. While the use of purely statistical methods to establish universals is generally frowned upon in linguistics today, the search for dictionaries of translatable semantic primitives has continued in the natural semantic metalanguage of Anna Wierzbicka (2021), and has recently been applied to knowledge organization (Hjørland 2026).

For the creation of the Horapollo and Real Character Language ontologies, I used the CLLD Concepticon dataset as a data source to organize my information (List 2016). The CLLD Concepticon is a resource from the world of linguistic linked open data (LLOD) which combines linguistics and ontology to create a federated meta KOS of over 400 linguistic concepticons. In this sense, it is a concepticon of concepticons, and represents a systematic appraisal of the latest research on the construction of such resources.

To derive the Horapollo Ontology, I coded each concept set in the CLLD concepticon using a qualitative coding scheme based on the criteria laid out in Kausch (2024).

2.3 Design of Icons

The name Horapollo is a reference to an encyclopedia of Egyptian hieroglyphs, published by the priest Horapollo Niliacus in the 5th century C.E. The encyclopedia presents allegorical explanations of both real and imagined Egyptian hieroglyphs, standing at odds with, or at times dubiously corroborating Egyptological interpretations of the phonetic signs (Iversen 1993). The ambiguity of the Hieroglyphica is that hieroglyphs are interpreted as symbols, when in fact they constitute a phonetic rebus for words in the Egyptian language.

To create images for the font, following the coding scheme which established which concept sets were most representable as images, I mapped the concept sets to specific characters based on the hierarchical arrangement of phonemes in linguistics called the

sonority hierarchy. Then I created a series of images for each concept set loosely inspired by the aesthetic of the reception of Horapollo's hieroglyphica; for this reason the initial vocabulary is called the Horapollo Ontology.

2.4 Prototype Theoretic Indexing

The application of the Real Character Language to a given vector database requires a mapping which can make each class have a meaningful relationship to the instances which are to be classified. Rather than using a discrete binary logic of classification where each instance either is or is not a member of a class, the goal is to use a fuzzy logic indexing technique, where each instance is a member of a class on a spectrum between 0 and 1. This has the added benefit of allowing every instance to be a member of every class to some degree. This produces a matrix where each class is a column, and the instances are rows, and each cell is a membership relation between 0 and 1; unlike with word embedding vector spaces produced by machine learning models, here each dimension is interpretable.

However, in order to classify a vector space for each new language or domain, there is a need to be able to determine the potential membership for all instances in a dataset, when this value is not known in advance. In order to compute the set membership, an application of prototype theory is used (Rosch 1973). A set of exemplar entities is selected from the vector space for each class. Then the cosine similarity from the average of these exemplars to every other item in the space is taken and the arccosine serves to map each item to a unit circle. This allows fuzzy membership to be computed as a proportion of distance rotated across the whole circle. This approach is easily generalizable to several domains and simple to visualize. It has issues in that some information is lost, and cosine similarity is an imperfect approximation of relatedness.

Another affordance this lends is the creation of wheel-based interfaces, also called volvelles (Schmidt 2017), based on the Ars Magna of Ramon Llull (Bonner 2007). When terms are placed on wheels, they can be rotated to adjust the context within the underlying vector space, making the system context-sensitive. There also is a "serendipity slider" which adjusts the focus from local to global. Higher serendipity takes into account the position of every term in the ontology based on their revolution about the circle, whereas lower serendipity only focuses on real characters on which the user focusses.

3.0 Conclusions

This paper describes exploratory steps towards a KOS which attempts several things. In addition to using fuzzy logic and exemplar theory to classify documents, it also explores how to unite vector databases and graph databases under a single KOS. This could potentially help make machine learning systems more explainable. It also is oriented towards multilingual document sets, and has a unique user interface which allows users to explore geometric and semantic transformations of the document space. Future work will investigate how more seamless data integration between RDF semantics and vector

databases can be achieved. Also, future studies involving user tests between multilingual participants to assess bias will be necessary.

Statement on the Use of AI: No AI was used to generate any part of this paper. Some AI was used for code generation for the website realcharacterlanguage.world.

References

Bacon, Francis. [190?]. The Works of Francis Bacon. Edited by James Spedding, Robert Leslie Ellis, and Douglas Denon Heath. Boston: Houghton, Mifflin. http://archive.org/details/worksfrancisbaco02bacoiala.

Bonner, Anthony. 2007. The Art and Logic of Ramon Llull: A User's Guide. Studien und Texte zur Geistesgeschichte des Mittelalters 95. Leiden and Boston: Brill.

Gnoli, Claudio. 2018. "Notation." Knowledge Organization 45: 667–684. https://doi.org/10.5771/0943-7444-2018-8-667.

Gnoli, Claudio. 2025. "International Society for Knowledge Organization." International Society for Knowledge Organization. Accessed June 3, 2025. https://www.isko.org/.

Hjørland, Birger. 2008. "What Is Knowledge Organization (KO)?" Knowledge Organization 35: 86–101. https://doi.org/10.5771/0943-7444-2008-2-3-86.

Hjørland, Birger. 2026. "Semantic Primitives and Compositionality: An Annual Review of Information Science and Technology (ARIST) Paper." Journal of the Association for Information Science and Technology 77: 198–223. https://doi.org/10.1002/asi.70011.

Iversen, Erik. 1993. The Myth of Egypt and Its Hieroglyphs in European Tradition. Princeton, NJ: Princeton University Press.

Kausch, John. 2024. "Nuclear Semiotics and Knowledge Organization: Five Design Heuristics for Semantic Primitives." In Knowledge Organization for Resilience in Times of Crisis: Challenges and Opportunities: Proceedings of the Eighteenth International ISKO Conference, 2024, Wuhan, China, edited by Wei Lu, Thiago Henrique Bragato Barros, Lu An, and Daniel Martínez-Ávila, 385–392. Advances in Knowledge Organization 20. Baden-Baden: Ergon. https://doi.org/10.5771/9783987400476-385.

List, Johann-Mattis, Michael Cysouw, and Robert Forkel. 2016. "Concepticon: A Resource for the Linking of Concept Lists." In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, et al., 2393–2400. Portorož, Slovenia: European Language Resources Association. https://aclanthology.org/L16-1379/.

Loveless, Natalie S. 2015. "Towards a Manifesto on Research-Creation." RACAR: Revue d'art canadienne / Canadian Art Review 40: 52–54. https://doi.org/10.7202/1032754ar.

Loveless, Natalie. 2019. How to Make Art at the End of the World: A Manifesto for Research-Creation. Durham, NC: Duke University Press. https://doi.org/10.1215/9781478004646.

Manning, Erin. 2016. "Ten Propositions for Research-Creation." In Collaboration in Performance Practice: Premises, Workings and Failures, edited by Noyale Colin and Stefanie Sachsenmaier, 133–141. London: Palgrave Macmillan. https://doi.org/10.1057/9781137462466_7.

Mazzocchi, Fulvio. 2018. "Knowledge Organization System (KOS): An Introductory Critical Account." Knowledge Organization 45: 54–78. https://doi.org/10.5771/0943-7444-2018-1-54.

McCormack, Derek P. 2008. "Thinking-Spaces for Research-Creation." Inflexions 1: 1–16.

Poornima, Shakthi, and Jeff Good. 2010. "Modeling and Encoding Traditional Wordlists for Machine Applications." In Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground, edited by Fei Xia, William Lewis, and Lori Levin, 1–9. Uppsala, Sweden: Association for Computational Linguistics. https://aclanthology.org/W10-2101/.

Rosch, Eleanor H. 1973. "On the Internal Structure of Perceptual and Semantic Categories." In Cognitive Development and the Acquisition of Language, edited by Timothy E. Moore, 111–144. New York: Academic Press. https://doi.org/10.1016/B978-0-12-505850-6.50010-4.

Schmidt, Suzanne Karr. 2017. Interactive and Sculptural Printmaking in the Renaissance. Brill's Studies in Intellectual History 270/21. Leiden and Boston: Brill. https://doi.org/10.1163/9789004354135.

Swadesh, Morris. 1955. "Towards Greater Accuracy in Lexicostatistic Dating." International Journal of American Linguistics 21: 121–137. https://doi.org/10.1086/464321.

Tennis, Joseph T. 2008. "Epistemology, Theory, and Methodology in Knowledge Organization: Toward a Classification, Metatheory, and Research Framework." Knowledge Organization 35: 102–112. https://doi.org/10.5771/0943-7444-2008-2-3-102.

Wierzbicka, Anna. 2021. "'Semantic Primitives', Fifty Years Later." Russian Journal of Linguistics 25: 317–342. https://doi.org/10.22363/2687-0088-2021-25-2-317-342.

Wilkins, John. (1668) 1968. An Essay towards a Real Character, and a Philosophical Language. English Linguistics 1500–1800, no. 119. Menston: Scolar Press.

Zadeh, Lotfi A. 1988. "Fuzzy Logic." Computer 21, no. 4: 83–93. https://doi.org/10.1109/2.53.

Zeng, Marcia Lei. 2008. "Knowledge Organization Systems (KOS)." Knowledge Organization 35: 160–182.