Larissa Terranova

Centro Studi Americanistici "Circolo Amerindiano", Perugia, Italy

From Stone to Signal: Multimodal Retrieval as Method in Archaeological Research

Archaeological knowledge lives in many forms: arguments, diagrams, maps, images, and the traces left behind. These forms do not exist in isolation; each one shapes and is shaped by the others as meaning is made. Where Digital Humanities has often centred on preserving and indexing, new multimodal systems open up possibilities for how we move through and interpret the tangled webs of complex archives.

This paper presents a multimodal retrieval architecture developed within the context of a doctoral corpus on Olmec urbanisation, part of the international *Ruta de la Obsidiana* research initiative. The project explores how semantic text retrieval, CLIP-based image alignment, and grounded generation can be integrated into a structured access layer over a curated archaeological archive.

The system brings together vector-based text retrieval, cross-modal similarity, diversity-aware ranking, and page-level metadata alignment to keep images and texts in meaningful proximity. An evaluation layer ("LLM-as-a-judge") checks for fidelity and relevance, thereby supporting traceability and transparency in response generation. By separating parametric and non-parametric memory, the archive remains open to updates and accountable to its sources. These choices shape how evidence is encountered, related to and made meaningful within the archive; they are not just technical. In practice, the system consistently retrieved page-aligned text and image pairs across languages, with groundedness scores reinforcing the traceability of each citation.

At its heart, archaeology is a discipline of attention. It is about learning to notice, to read meaning in fragments, gaps, and the ways things relate. It trains the researcher to read traces, recognise the meaning distributed across fragments, interpret absences, and discover relationships across material. In this sense, the design of computational systems for archaeology is never neutral. It encodes assumptions about what counts as proximity, relevance, and connection. By structuring retrieval across text and image simultaneously, this project treats technological architecture as an extension of interpretative practice. The system does not replace the slow work of analysis; it reorganises the conditions under which that work becomes possible.

In this field, interpretation grows from relationships: artefact to landscape, image to inscription, fragment to structure. Retrieval is not just a technical step but a way of shaping how these relationships are traced and understood. A multimodal system that honours these connections can help researchers move through archives without losing sight of the evidence that grounds their work.

This work situates the prototype within broader Digital Humanities discussions on sustainability, epistemic responsibility, and infrastructural design. Multimodal retrieval is seen here as a way to open access while preserving the archive's complexity. The goal is not just to speed things up, but to align technology with the ways archaeological knowledge is built and shared.

If archives hold memory, retrieval systems shape how we meet it. Thoughtful design is part of how we care for heritage today.