

Analysis of Cultural Heritage in the Age of Artificial Intelligence

Experiences from Mexico

Diego Jiménez-Badillo

National Institute of Anthropology and History diego_jimenez@inah.gob.mx

7th. International Workshop on Analysis, Understanding and Promotion of Heritage Contents, October 27th 2025, Dublin.

Artificial Intelligence

Empowers communities

- Automatic translation of endangered languages
- Citizen monitoring of cultural heritage

Facilitates access to CH collections

- Discover objects and establish meaningful connections
- Automatic cataloguing of objects

Improves conservation

- Automatic reassembly of pieces
- Virtual reconstruction
- Monitoring of deformations and structural changes in artworks

Makes communication in museums more meaningful

• Personalization of content for specific users by analysing location, weather, news, historical data, etc.





Unleashing the potential of Machine Learning

Text analysis

Data retrieval

Shape Analysis

- Automatic classification
- Style attribution
- Determination of provenance
- Answering questions such as:
- How the shape of an object relates to its function
- What concepts, symbols or ideas are conveyed through the shape of an object, symbols odd



1

Analysis of historical documents with NLP techniques

Project Overview

This research project applies computational linguistics and machine learning techniques to facilitate the analysis of the *Relaciones Geográficas de la Nueva España*. A corpus of reports compiled between 1577 and 1585.

Methodology

The approach integrates:

- Named Entity Recognition and Disambiguation (NERD)
- Geographical Text Analysis

The primary goal is to assist scholars from various disciplines in efficiently extracting structured information from historical documents.

Project objective

The primary objective of this project is to provide scholars across diverse disciplines with advanced tools to efficiently retrieve relevant information, integrate it with complementary documentary resources, and address significant research questions, such as:

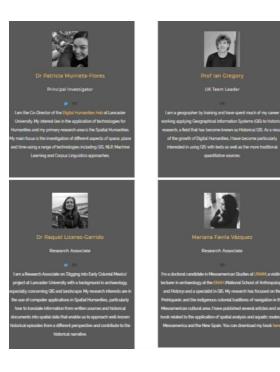
- Where did certain phenomena occur most frequently?
- Who was responsible for specific events?
- What transpired in particular locations at given points in time?

Impact

By enabling systematic access to interconnected historical and documentary data, the project fosters more rigorous and comprehensive analyses of complex phenomena. Researchers are able to uncover spatial and temporal patterns, establish causal relationships, and generate new insights that were previously inaccessible due to fragmented or siloed information. Ultimately, this initiative is expected to advance interdisciplinary scholarship, support evidence-based historical interpretations, and contribute to the development of innovative methodologies for data-driven research in the humanities and social sciences.

Research group

- Lancaster University, UK
- National Institute of Anthropology and History. Mexico
- Lisbon University, Portugal







- Patricia Murrieta-Flores
- Diego Jiménez Badillo
- Bruno Martins
- Raquel Liceras Garrido
- Mariana Favila Vázquez
- Katherine Bellamy
- Luis P. Santos

General questions

- What topics do the texts address?
- Which locations are mentioned?
- What is said about each location?
- Which individuals are associated with specific events, and where did these events occur?
- Who was responsible for particular actions?
- What happened in specific places at specific times?
- What relationships exist among the entities mentioned in the texts (e.g., connections between individuals, associations between people and places or events)?

The Relaciones Geográficas from New Spain (1577-1585)

The corpus consists of responses to a questionnaire commissioned by King Philip II, designed to gather information about the newly acquired territories in New Spain.

The survey included fifty questions and provides insights into:

- History
- Geography
- Administrative jurisdictions
- Culture and religion
- Economy
- Social interactions
- Processes of transculturation affecting both native communities and colonizers



King Philip II

Instructió, y memoria, de las relaciones que

fehan dehazer,para la descripcion de las Indias, que su Magestad man dahazer,para el buen gonierno y ennoblefcimiento dellas,



Rimeramente,los Gouernadores, Corregidores, o Alcaldes mayores, a quien los Vireyes, o Audiccias, y otras plonas del gouier no,embiaren estas inftructiones, y memorras impreifas, antecodas cofas haran lifta, y memoria de los pueblos de Españoles, y de Indionque vuiere en fujurifdittion, en que folamente seponga los nombres de ellos escriptos de letra legible, y clara, y luego la embiaran a

las dichas personas del gouserno, para que juntamente, con las relaciones que en los dichos pueblos se hizieren la embien a la Magestad, y al consejo de

Y diftribuyran las dichas instructiones, y memorias impressas por los pueblos de los Españoles, y de Indios, de su jurisdició, donde vuiere Españoles, embiando las alos concejos, y donde no, alos curas filos vuiere, y fino alos religio sos, a cuyo cargo suerela do arina, mandando a los concejos, y encargando de partedela Magestad, alos curas y religiosos, que dentro de va breue termipartesetu ingentasi ato cura y reingosto, duciento even brene termi-no, las refpondano, fatiriagan como enclias dedeciara, y les embienlas rela-ciones que hizieren, antamétecon ellas memorias, para que ellos como fue ten recibiendo las relaciones, vayan embiando las alas perionas de gouierno que felas vuieren embiado, y las inflirucitiones y memorias las buelvan a dif tribuy rificare membre por los otros puchios adode no las vuiere embiado. Y en los pueblos, y ciudades, didelos Gouernadores, o Corregidores, y perfo-

nas de gouierno relidieren, haran las relaciones de ellos, o encargar las han a personas intelligentes de las cosas de la tierra que las hagan, segun el tenor de

Las personas a quien se diere cargo en los pueblos de hazer la relacion parricular de cada vno dellos, responderan a los capitulos de la memoria, que se

figue por la orden, y forma figuiente.

Primeramente, en vn papel a parte, pódran por caueçade la relacion que hizie-ren, el dia, mes, y año de la fecha de el la com el nombre de la persona, o perso. nas que se hallaren a hazeria, y el del Gonernador, votra persona que les vuie

reembiado la dicha instruction,

reembiado la dicha infiredition.

Yleyendoarentamente, rada Capitulo dela memoria, foreairálo que humere que dezir a elen otro capitulo por fix elegondiendo a cada vno por fius muneros, como van en la memoria, vno trasotro y en los que no humere que dezir, de zar los há fin hazer mécion de ellos, y pafíaran a los figulentes, halha acauar los delecertodos, y tefponder los quuieren que dezir, como queda dicho, bre se y claramente, entodos afirmando por cierto lo que lo fuere, y lo que no, poniendolo por dadofo demanera que las relaciones venganciertas, conforme a lo contenido en los capitulos figuientes.

Historical Context of the Questionnaire

- 1553–1557: Alonso de Santa Cruz initiates early efforts
- **1569**: Juan de Ovando sends a 37-question survey
- **1570**: A new survey with 200 questions
- **1573**: A third version with 135 questions
- **1577**: Juan López de Velasco condenses the questionnaire to 50 questions, forming the basis of the current corpus
- Responses were received from New Spain and Guatemala between 1578 and 1585

Sample Questions and Responses

Demographics and Settlement Patterns

Question 5: "Is the town populated by many or few Indigenous people? Has the population changed over time, and why? Are they settled in permanent towns? What are their intellectual and cultural traits, inclinations, and ways of life? Are there multiple languages spoken in the province, or is there a common language?"

Response: "This head town has 410 married tributary Indigenous people. It had more in the past, but the number has decreased due to pestilence. The current population lives in permanent settlements. Their intellectual and cultural traits are limited and lack refinement. Their way of life includes efforts to bring them into Christian doctrine and their drinking habits. They all speak one language: Mexican." (Relación de Quautlatlauca y Huehuetlan)

Natural Resources

Question 22: "What wild trees are commonly found in the region, and what fruits or benefits do they provide? What are they good for?"

Response: "Wild trees include large ones called *izamatl* by the Indigenous people, meaning 'razor and paper,' as well as thorny trees and others called *guaxes*. Fruit-bearing and useful trees include custard apples, lemons, guavas, citrons, oranges, magueys, and prickly pears. Spanish trees that grow here include quince, pomegranate, and grapevine."

Agricultural Products

Question 24: "What grains, seeds, vegetables, and greens serve or have served as sustenance for the natives?"

Response: "The seeds grown in this region include all kinds of vegetables and greens, wheat, chickpeas, native peppers, and pumpkins from both Castile and the local area." (*Relación de Quautlatlauca y Huehuetlan*)

Medicinal Plants

Question 26: "What aromatic herbs or plants are used by the Indigenous people for healing, and what are their medicinal or poisonous properties?"

Response: "The herbs used for healing include *coanenepilli*, meaning 'snake tongue,' which is given to those with fevers and dysentery; it is very refreshing. Another root, *tlaizotlalti*, means 'that which expels what is in the body,' and is very beneficial."

Fauna

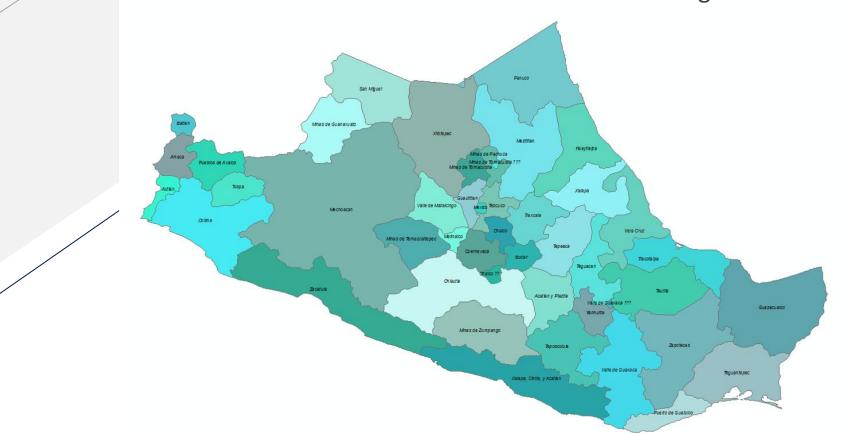
Question 27: "What wild and domestic animals and birds exist in the region, and which ones were brought from Spain? How do they breed and multiply?"

Response: "Wild animals include large tawny lions, coyotes (called *adives*), wild cats, and deer. Birds include hawks, falcons, owls, crows, and small deer. There are native wild hens and tame hens from both Castile and the region, as well as pigs."

- 168 text reports (around 1,000,000 words)
- 78 pictographic maps

Geographic coverage

- 50 provinces
- 168 main jurisdictions (cabeceras)
- 248 secondary jurisdiction(*sujetos*)
- 414 main towns
- More than 2000 villages and small towns



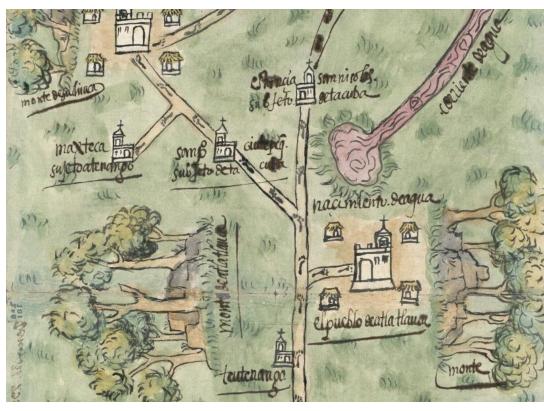
Relación Geográfica from Zempoala, Hidalgo

sen pu Juala nel qui ce senquagnala sela niceba conquegación sela puebe españo en grim se nobientre se mill y que y ocenta anot lelita full des obregon or negative se mil good y ordent and feel quelle se pasaguer y historian substitute of puelle se pasaguer y historian sur palear par for me to farecer such si a on a se mender prentis pal se fengualisat que su fam se guerma potennan sel que saquala yaon palle se aquato sobernado se tecquipo ya min solució potentado sel se anyly rales aller) select stops puebles ya to the mager ynder tothe y ancianos selve reso que ellos que sa Jor la soccupación que su moj maiso selve reso que elos qui intercente fu se numa a exponen sela longua mexicana interprete fucion que guntado par los ores the bol selver sescrepain foxeron yallatarion & agmentes . Cos quellos sela congregación se genquala son quato cabeceras lamase jen sua Suata Raguel po Saquala tecquesa de verging ella mita se thaquiton d'inn tos e for quart preches funter per souse le llaman son fie gauen. Fran un lland herras se faquela gentrafia a con gregaron se pa chor Junar ala dottima y per causa sel aqua que antes nota tenjon fino saguezes) furcion Antiga aqua Ima legua de sinde Ann songregado del que Se on corto glaman He castitlan que que usefor cof rijod maneta se fragers) pa que onla cubic selones ceffe for on lland Elfo some huffer theyeren chaques poronol arcord de cal years fafa en meles dela on quegación en ma fuente en m - och place odst gnaff rigst puetter) y some port das les alles thapes franksor auguegarouse d'an se mill y grier fin grente sete sa se san min penquala tiches quat fuseat que se llaman figts nafuacthepa. que afua afua quaughtanthe quilpo tiene tes suferor que son alteran contexes nopalapor taquala here grant fulcos que genten ac xotta nexteno the te co mulos sary to cer teo fel partiene grant fulcor teal negga tetsaguagan nequa meterer fagallon en 4016 chos sufers tienen Thepas quelos aptom Go orcher Hagres films of acuser ala softime about rigor quellos ortim solar enternat alequa yalequaymega yas Cequas TIN Aut Tregos quart quebles from ofe leguar sela flow se me dance for los quellos es marcanos que trene son pagaca quello co con les porques aquat (egnas) afen el nive te Aunha



Relación Geográfica from Atlatlauhcan (Morelos)





Methodology

1. Corpus Preprocessing

- Texts were digitized using Optical Character Recognition (OCR), enabling machine-readable formats.
- Content was normalized to ensure structural and linguistic consistency for computational processing.

2. Named Entity Recognition and Classification

- Tokenization segmented the text into basic units (words, punctuation, numbers, dates).
- Initial entity identification relied on syntactic roles, token positions, contextual clues, and lexical resources such as specialized dictionaries and a custom gazetteer.
- Entities were classified according to a custom ontological schema including categories such as individuals, institutions, flora, fauna, cultural elements, religious rituals, and natural resources.
- SpyC algorithms were used to train the recognition model.

3. Disambiguation

- Linguistic and geographic ambiguities were addressed through supervised analysis.
- Special attention was given to toponym recognition, essential for georeferencing historical, geographic, and cultural data.
- Each placename was assigned latitude and longitude coordinates.
- A specialized directory of 16th-century toponyms was compiled.

4. Contextual Entity Analysis

 A word embedding algorithm was applied to disambiguate entity meanings based on context. For instance, "San Francisco" could refer to a person, a city, or a religious institution depending on its linguistic environment.



conforme a la dicha Instrucción, así el dicho Cabildo como el dicho Prior, hiciesen las Relaciones conforme a los capítulos de las dichas Instrucciones, y su merced del dicho señor Alcalde Mayor tornó a su cargo los demás pueblos, que es la Chontalpa, para hacer la dicha Relación y les apercibió por sus cartas con brevedad lo despachasen, que en lo que le tocaba haría lo mismo, para que hecho se despachasen y enviase al dicho señor Gobernador, para que su merced lo enviase a Su Majestad y su real Consejo de las Indias, y así lo proveyó y firmó de su nombre, que fue fecho el dicho día,

142 <

Food

maíz

1142 ≠ 0 (0.00%) IAA: -

Towns mentioned in the Relaciones Geográficas



Map of the 16th-century towns geo-located during the course of the project. The 46 polygons painted in different colors delimit some districts (*alcaldías mayores*) of the Government of New Spain and the Government of Yucatán, roughly following the work of Gerhard (2000: map 7).

Project outputs

As part of our initiative to enhance access to the *Relaciones Geográficas* corpus, we have developed two online tools designed to facilitate the extraction and analysis of historical and geographical information.

1. Web Platform for the Relaciones Geográficas

The first tool is a comprehensive web-based platform that, for the first time, consolidates all available data related to the *Relaciones Geográficas* corpus. It is intended for users who wish to explore the corpus without engaging in complex search procedures and who seek access to diverse forms of information, including:

- High-resolution images of the original manuscript folios
- Various published transcriptions of the documents
- Interactive maps displaying the location of all toponyms mentioned in each report
- Pictographic maps associated with the corpus

The platform is accessible at:

https://relacionesgeograficas.inah.gob.mx/



Este sitio reúne imágenes, transcripciones, mapas e información temática del conjunto de documentos conocido como Relaciones Geográficas de la Nueva España, específicamente los reportes redactados entre 1577 y 1585 en más de la mitad de las provincias de México y Guatemala por orden del rey Felipe

La organización de estos informes sigue un cuestionario diseñado por Juan López de Velasco, cosmógrafo real y miembro del Consejo de Indias, y describen cómo eran la organización territorial y el modo de vida de los habitantes de Nueva España, seis décadas después de consumada la conquista de México-Tenochtitlan.

Las Relaciones se componen de textos que abarcan casi un millón de palabras y 78 mapas pictográficos, en cuya elaboración participaron informantes indígenas y oficiales españoles. Debido a la riqueza y variedad del contenido, el corpus constituye una de las fuentes más importantes para analizar la historia, la geografía, las jurisdicciones administrativas, la cultura, la religión, la economía, la interacción social y los procesos de transculturación que afectaron a las comunidades nativas y a los colonizadores europeos.





El objetivo de este sitio es fomentar el interés público en estos documentos y facilitar el trabajo de historiadores, etnólogos, arqueólogos y lingüistas interesados en el corpus. Cada relación geográfica se acompaña de un mapa topográfico que muestra la localización de los pueblos y los rasgos geográficos mencionados en los documentos. Además, en los casos de relaciones que incluyen un mapa pictográfico, éste se despliega en alta resolución para que todos sus detalles sean apreciados fácilmente. Asimismo, se presenta una imagen de cada foja junto a la transcripción correspondiente. Finalmente, se incluyen referencias bibliográficas de las publicaciones dedicadas a transcribir, editar o analizar cada relación geográfica.

La producción de este sitio cumple uno de los objetivos del proyecto Explorando el México Colonial Temprano. Análisis computacional a gran escala de fuentes históricas del siglo xvi. Y ha sido posible gracias a la participación de diversas instituciones que han otorgado los permisos para la difusión de estos materiales en línea. Particularmente, la Llilas Benson Latin American Studies and Collections y el Archivo General de Indias, que resquardan una buena parte de los documentos originales. así como el Instituto Nacional de Antropología e Historia (INAH) y el Consejo Nacional de Ciencia y Tecnología de México (CONACyT).



Puedes consultar las relaciones geográficas por medio de las dos opciones siguientes:



























Resumen

Este artículo describe el conjunto de documentos conocidos como Relaciones Geográficas de la Nueva España, específicamente los reportes redactados entre 1577 y 1585 en más de la mitad de las provincias de México y Guatemala por orden del rey Felipe II.

La organización de estos informes sique un cuestionario diseñado por Juan López de Velasco, cosmógrafo real y miembro del Consejo de Indias, y describen cómo eran la organización territorial y el modo de vida de los habitantes de Nueva España seis décadas después de consumada la conquista de México-Tenochtitlan.

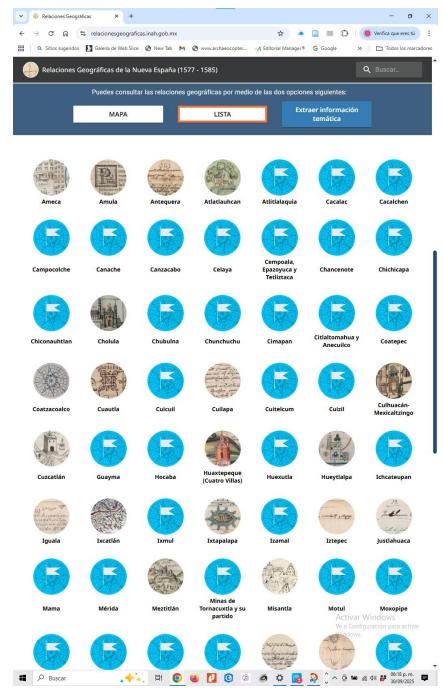
Las Relaciones se componen de textos que abarcan casi un millón de palabras y 78 mapas pictográficos en cuya elaboración participaron informantes indígenas y oficiales españoles. Debido a la riqueza y variedad del contenido, el corpus constituye una de las fuentes más importantes para analizar la historia, la geografía, las jurisdicciones administrativas, la cultura, la religión, la economía, la interacción social y los procesos de transculturación que afectaron a las comunidades nativas y a los colonizadores europeos.



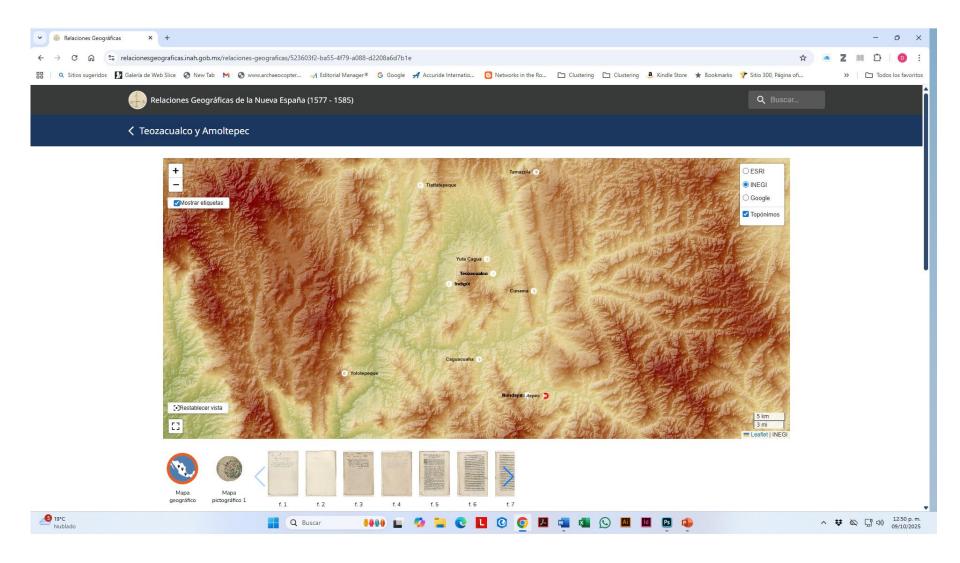


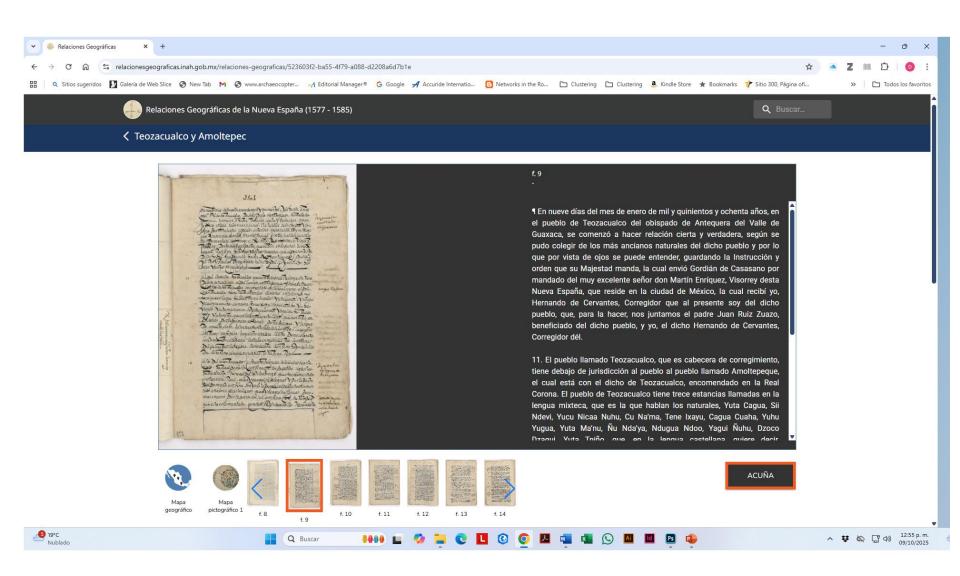


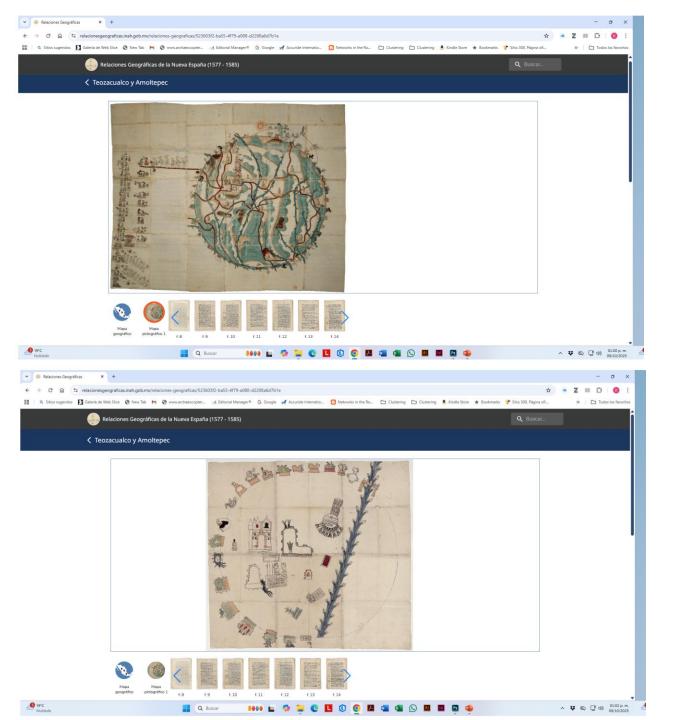


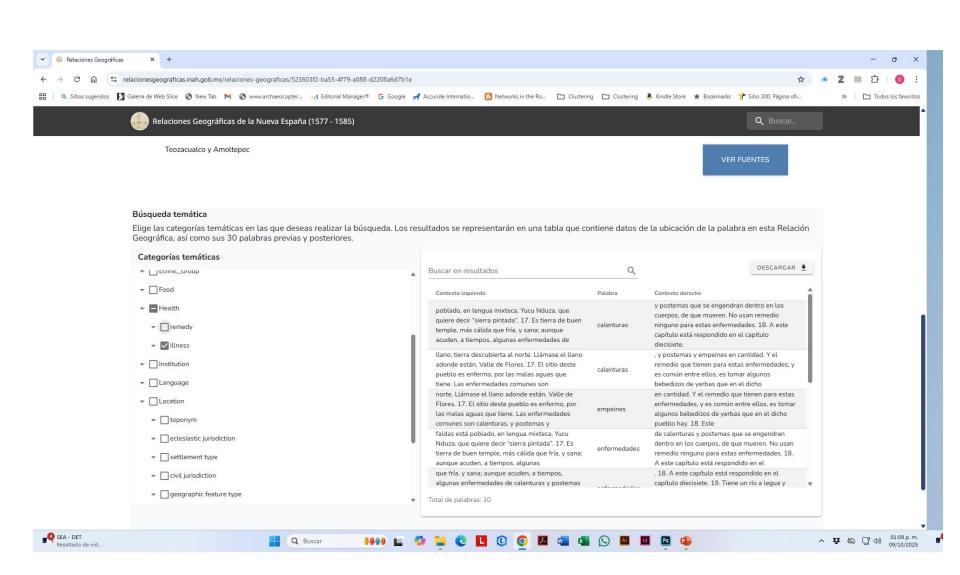


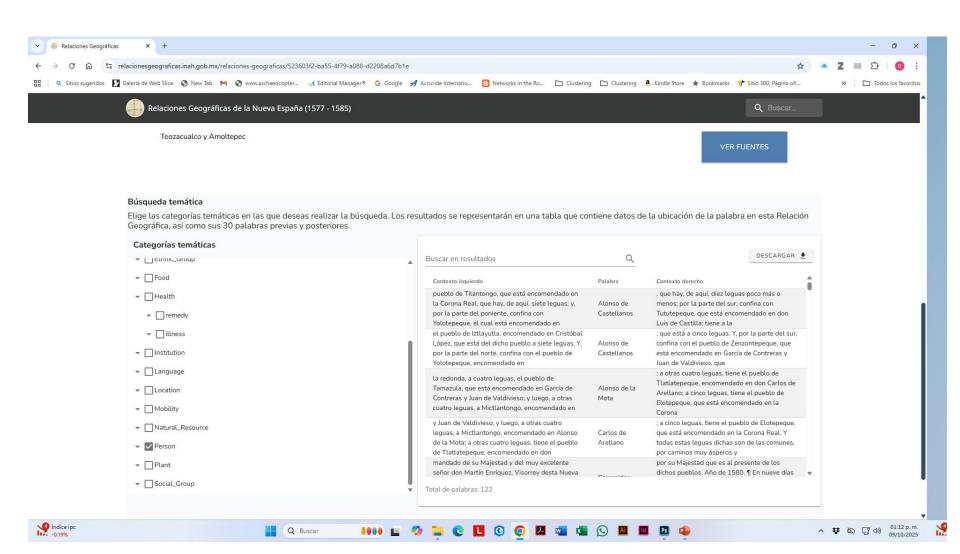
Relación Geográfica de Teozacoalco y Amoltepec

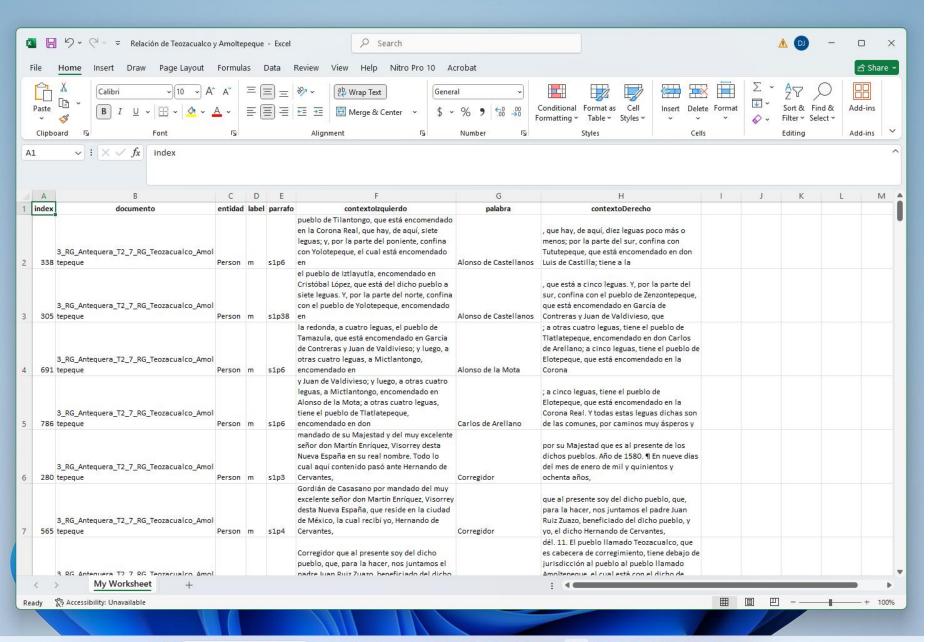






































Geographical Text Analysis

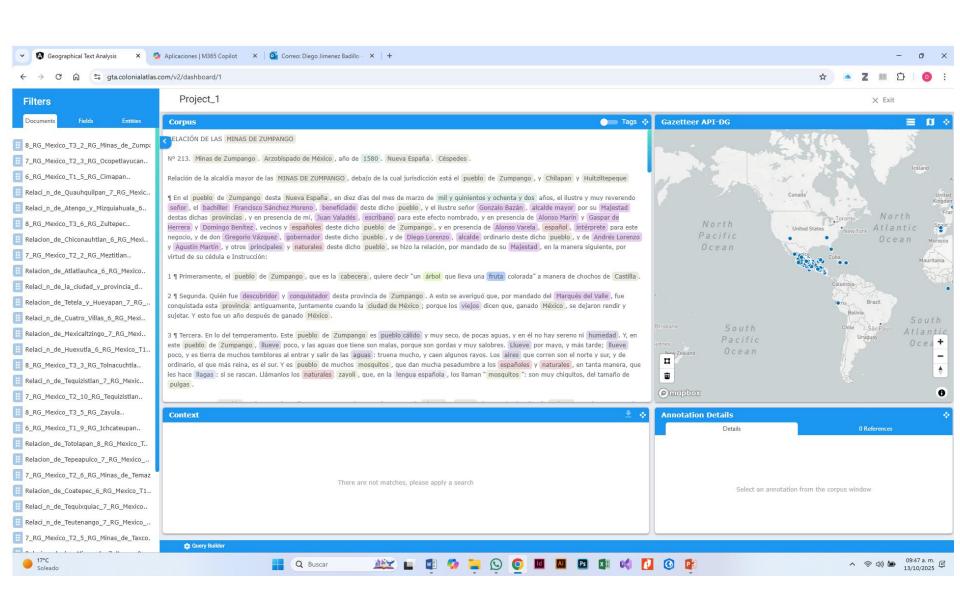
The second tool, named **GTA** (**Geographical Text Analysis**), enables researchers to perform customized queries based on their specific interests. It allows for the extraction of significant words or phrases from the corpus and facilitates contextual analysis.

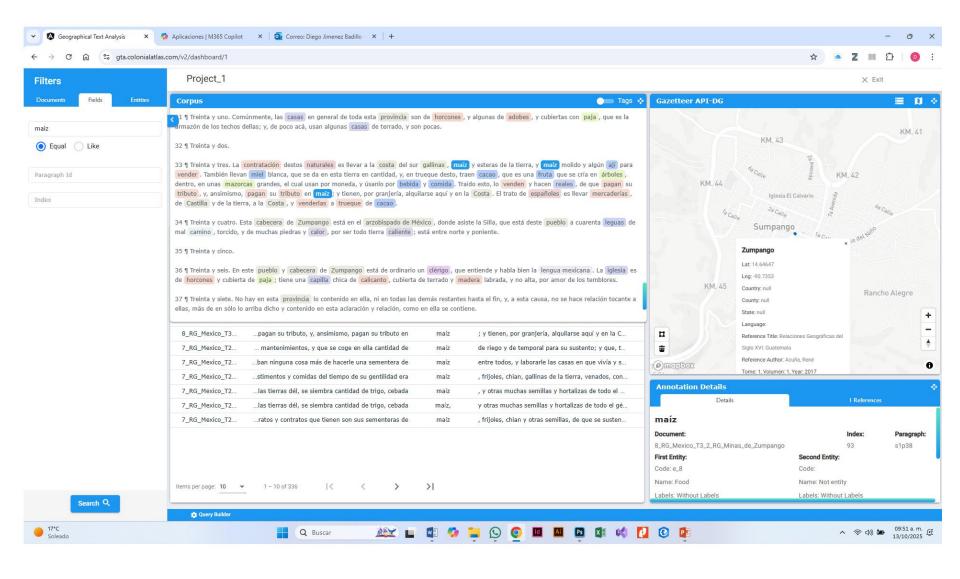
This tool supports research in areas such as:

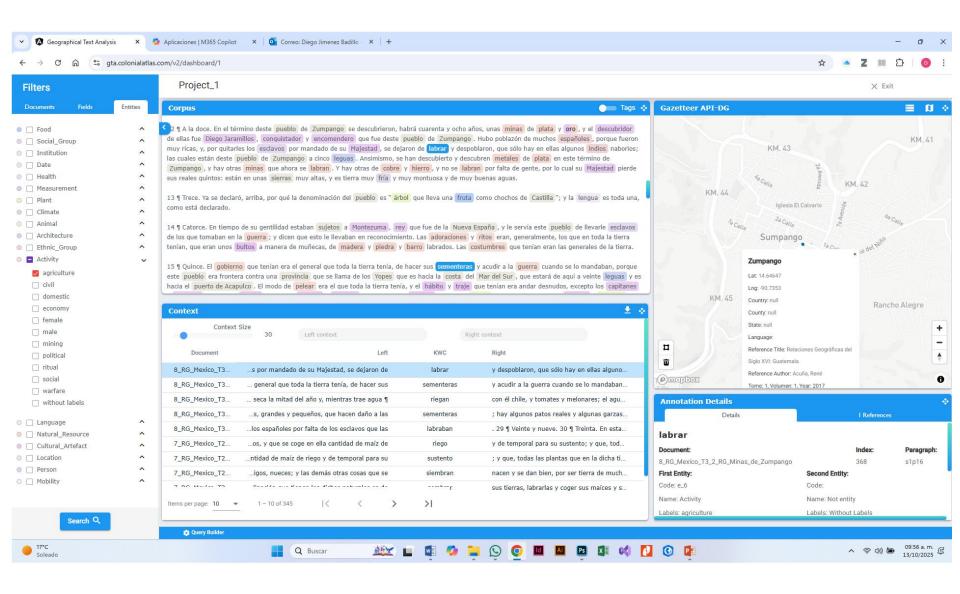
- Geographical correlation of historical events
- Analysis of social, cultural, and historical data linked to geo-referenced placenames
- Reconstruction of historical territorial jurisdictions using spatial information from the documents
- Identification of interaction patterns among individuals, institutions, activities, and locations
- Searches are conducted by entering a keyword of interest and defining a proximity threshold, typically based on the maximum number of words between the keyword and its surrounding context.

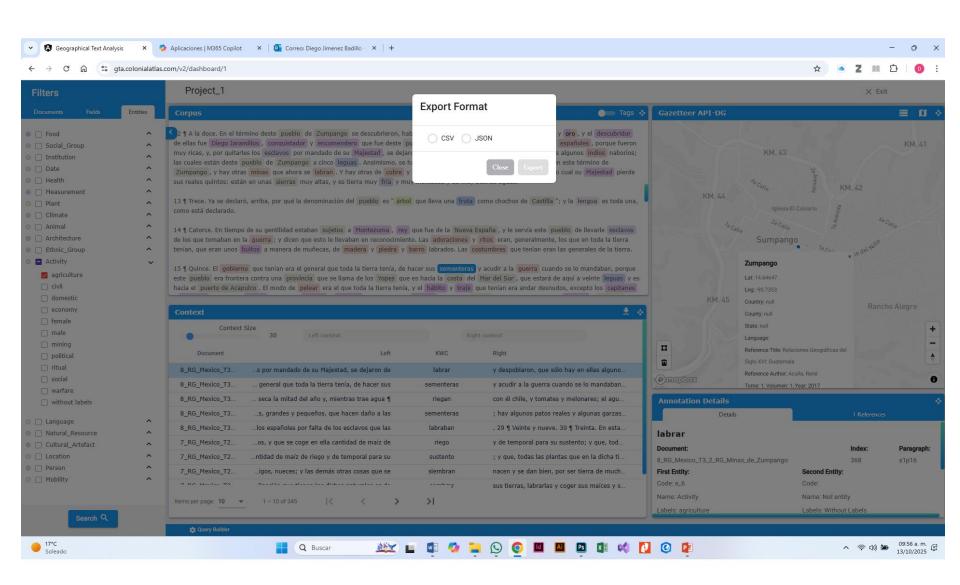
Access the tool at:

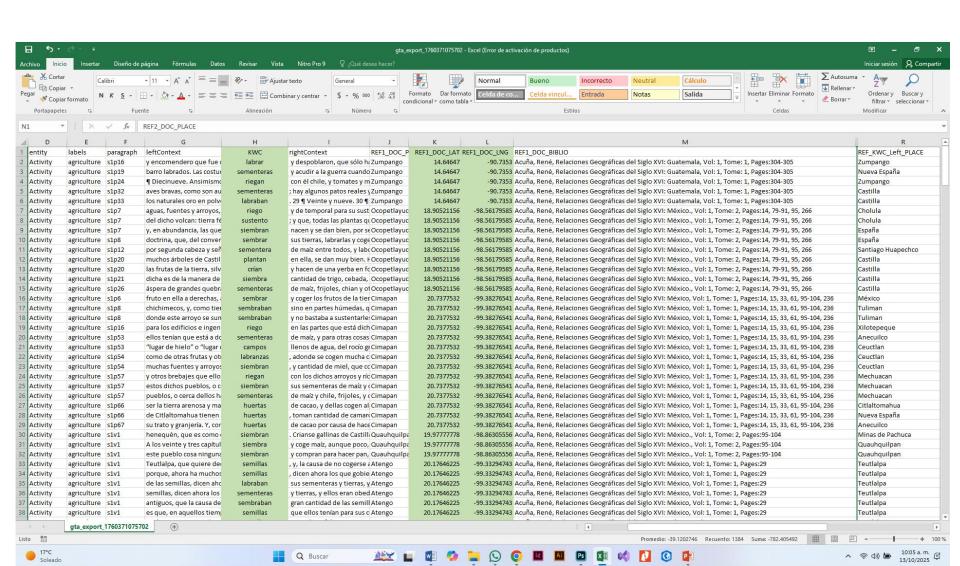
https://gta.colonialatlas.com/v2/





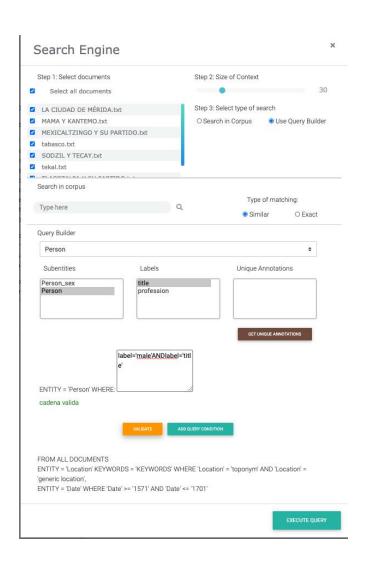






El módulo query builder de la herramienta GTA

Interface of the tool for building complex queries (Query Builder), which allows combining search criteria using Boolean operators such as AND, OR, XOR, etc.



Repository

GitHub: https://github.com/patymurrieta/Digging-into-Early-Colonial-Mexico/tree/master/DECM_Corpus/DECM_Machine_Annotated_Corpus

Analytical Capabilities

These tools enable the automatic identification and extraction of data, supporting quantitative and qualitative analysis.

The types of information that can be retrieved include:

- Prominent public figures, clergy, government officials, and individuals in positions of power during the 16th century
- Infrastructure related to religious orders, hospitals, churches, and other institutions
- Geographic and demographic descriptions of each town mentioned in the texts
- Linguistic data concerning the languages spoken across different regions
- Trade and exchange relationships among the communities referenced in the corpus
- Festivities, customs, religious beliefs, and political organization of Indigenous populations

A system for automatic recognition of 3D archaeological models

The objective is to facilitate the dissemination of cultural heritage objects through the implementation of a search engine that operates using 3D object recognition algorithms.

The system resembles "Google," but instead of using text for queries, it relies on 3D geometric information.

The Shaping Archaeology Group



Diego Jiménez-Badillo

INSTITUTO NACIONAL DE ANTROPOLOGÍA E HISTORIA (INAH)



Salvador Ruiz-Correa

Instituto potosino de investigación en ciencia y tecnología (ipicyt)



Omar Mendoza-Montoya

Instituto tecnológico de Monterrey



Edgar F. Román-Rangel

Instituto tecnológico autónomo de méxico (itam)



Mario Canul Ku and Rogelio Hasimoto Beltrán

CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS (CIMAT)







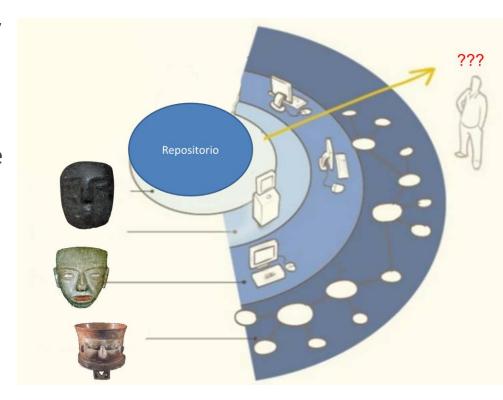
Current challenge

No easy access to digital 3D models

Digitization through photogrammetry and laser scanning has produced millions of digital models of cultural objects.

Unfortunately, once these models are stored in a digital repository both researchers and the general public face serious obstacles in accessing these resources online.

This is mainly due to the lack of appropriate query tools for retrieving 3D information..



Traditional query

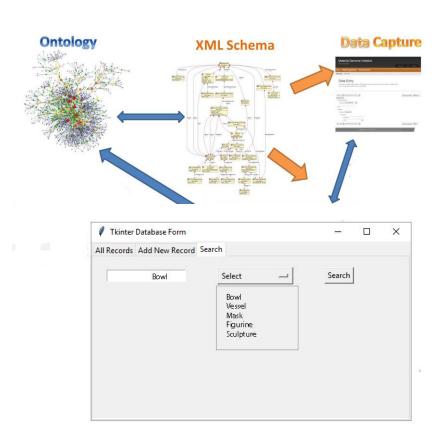
Use of keywords that describe the objects being searched for, for example:

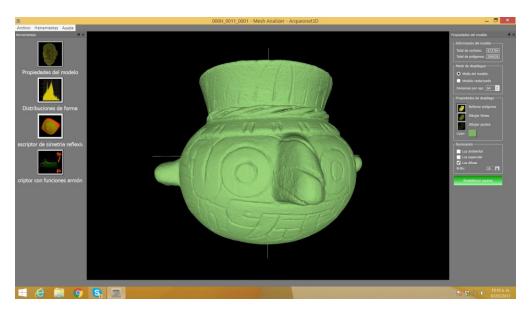
"retrieve all tripod bowls"

This method only works if the words used belong to the vocabulary recognized by a particular database, since it might recognize bowl but not cajete (a term more commonly used in Mesoamerican archaeology).

Another limiting factor is the language in which the descriptions have been written, as the same database might recognize the term cuenco (Spanish) but not the words bowl (English) or terrine (French).

Therefore, this traditional query method is inadequate for general-purpose, multilingual searches focused on analyzing the shape of objects.





SOLUTION

The solution is to develop a system that searches for objects within a 3D model repository by analysing only their geometric features—without relying on textual queries.

1. Query Input

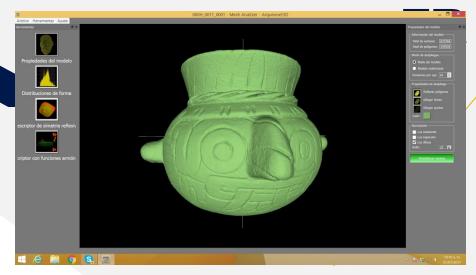
The user provides a query model, which can be a point cloud or a surface mesh representing the shape of the object they wish to find in the database—for example, a vessel like the one shown in the figure.

This step replaces the need for textual input such as "retrieve all masks with prominent noses"

2. Geometric Comparison

The system then compares the geometry of the query model against all 3D models stored in the repository.

The goal is to identify those models that resemble the query, regardless of whether the collection contains dozens, hundreds, or thousands of examples.



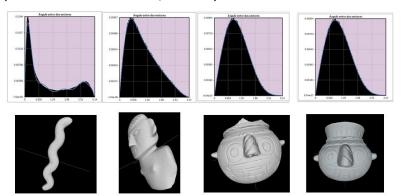


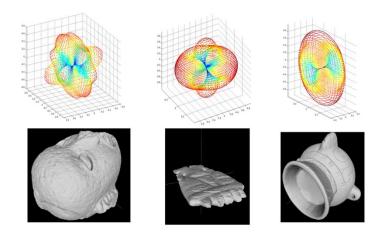
Add a footer 47

Shape descriptors

Distance measurements

(Osada et al. 2001, 2002)

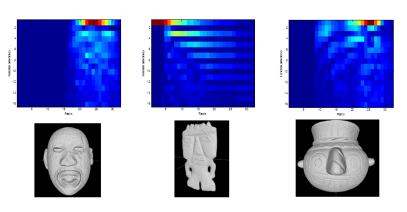


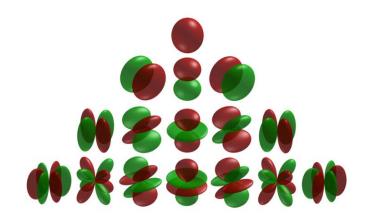


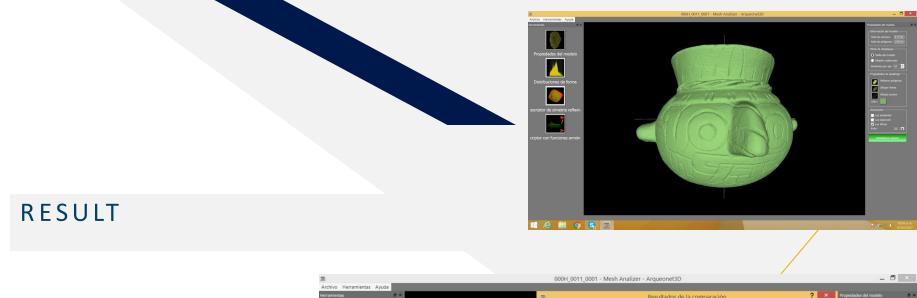
Reflexive symmetry (Kazhdan et al. 2002, 2004a, 2004b)

Harmonic functions

(Kazhdan and Funkhouser 2002; Kazhdan et al. 2003).



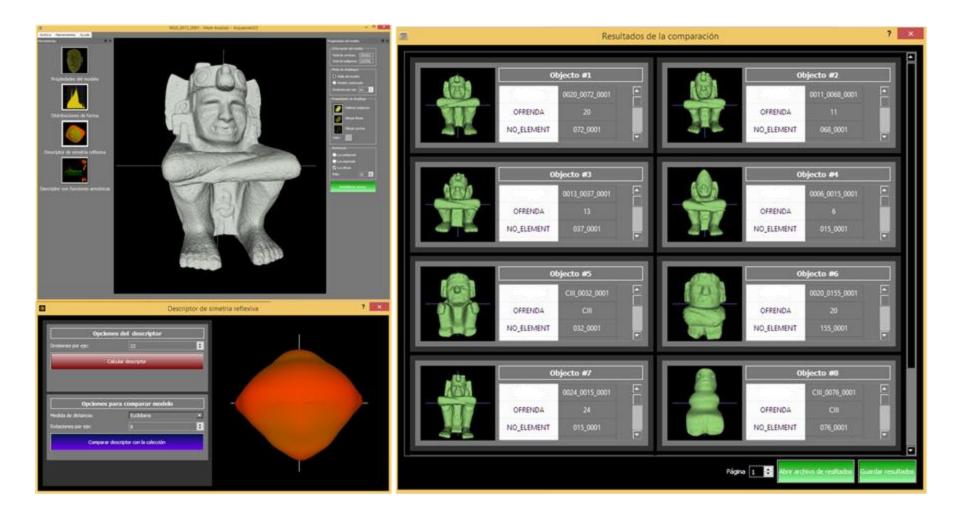


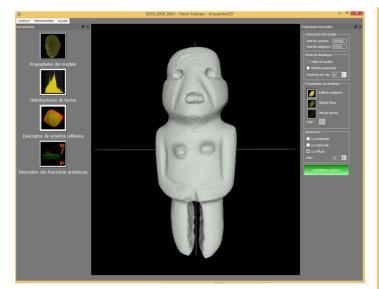


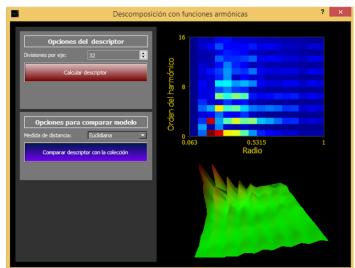
Finally, the system displays the results on screen, sorted from most to least similar to the query model.

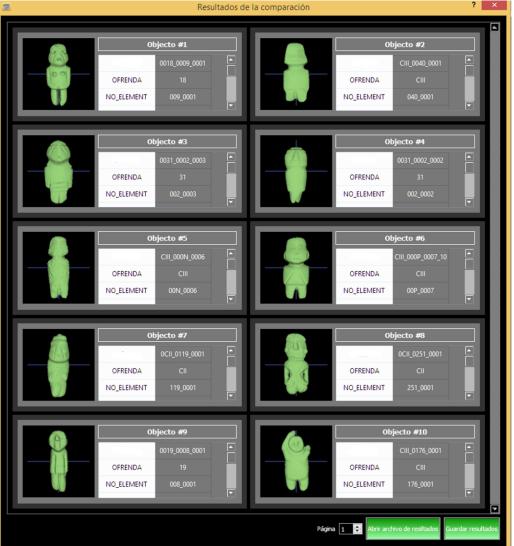


Query examples

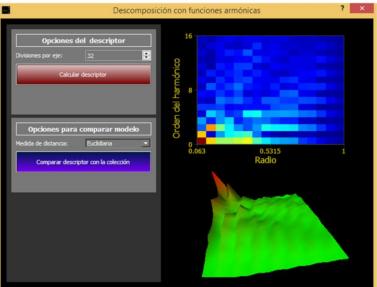














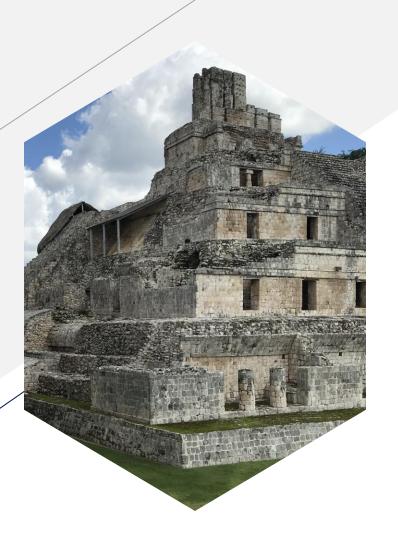
Software and sample data

- Software: DOI <u>10.5281/zenodo.7411318</u> (Mendoza-Montoya, 2023a).
- 3D models sample DOI <u>10.5281/zenodo.7583649</u> (Mendoza-Montoya, 2023b).

References

- Kazdhan, Michael, et al. "A Reflective Symmetry Descriptor". *Proceedings of the 7th European Conference on Computer Vision, ECCV02*, Copenhagen, May, 2002. *Lecture Notes on Computer Sciences*, vol. 2352 (Part II), pp. 642-656. https://doi.org/10.1007/3-540-47967-843.
- Kazdhan, Michael, et al. "Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors". *Proceedings of the* 2003 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing (SGP '03), 2003, pp. 156-164. https://dl.acm.org/doi/10.5555/882370.882392.
- Kazhdan, Michael, and Thomas Funkhouser. "Harmonic 3D Shape Matching". <u>SIGGRAPH '02: ACM SIGGRAPH 2002 Conference</u>

 Abstracts and Applications, 2002, p. 191. https://doi.org/10.1145/1242073.1242204.
- Kazhdan, Michael., et al. "A Reflective Symmetry Descriptor for 3D Models". *Algorithmica*, vol. 38, 2004a, pp. 201–225. https://doi.org/10.1007/s00453-003-1050-5.
- Kazhdan, Michael, et al. "Symmetry Descriptors and 3D Shape Matching. *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing (SGP '04)*, 2004b, pp. 115-123. https://doi.org/10.1145/1057432.1057448.
- Osada Robert, et al. "Matching 3D Models with Shape Distributions". *Proceedings of the International Conference on Shape Modelling and Applications*, Genova, Italy, 2001, pp. 154-166. DOI: 10.1109/SMA.2001.923386.
- Osada, Robert, et al. "Shape Distributions". ACM Transactions on Graphics, vol. 21, 2002, pp. 807-832.
- Shelton, C. R. (2000). Morphable Surface Models. International Journal of Computer Vision, 38(1): 75-91.



Thanks

- Diego Jiménez-Badillo
- +52 1 55 2193 2972
- ☑ diego_jimenez@inah.gob.mx
- www.inah.gob.mx