



XII CoLiCo

Coloquio de Lingüística Computacional

Con las ponencias magistrales de
Chandel Dávila Gutiérrez e Ismael Gómez García
Organización de Estados Iberoamericanos

Facultad de Filosofía y Letras
Grupo de Ingeniería Lingüística
Instituto de Investigaciones Bibliográficas
Escuela Nacional de Ciencias Forenses

Auditorio José María Vigil
de la Biblioteca Nacional de México
Centro Cultural Universitario, Ciudad Universitaria
Mas información: www.colico.unam.mx / colico@iingen.unam.mx

9-10
septiembre | 2025
10-19 h



Se otorgará constancia de asistencia

Transmisión
por
YouTube

Mesa 1. EXTRACCIÓN DE DATOS

Autor(es): Gerardo Sierra y Martha Romero

Institución: UNAM-IIB

Título: Extracción automática de metadatos de la Biblioteca Nacional de México



Resumen: La Biblioteca Nacional de México, como repositorio patrimonial de la memoria escrita del país, realiza actualmente la extracción de metadatos de manera manual, a cargo de catalogadores expertos, para apoyar la catalogación temática y bibliográfica de documentos pertenecientes al patrimonio cultural mexicano. Aunque este método garantiza alta precisión, resulta costoso en términos de tiempo y recursos humanos.

Con el objetivo de optimizar este proceso, se propone un sistema de extracción automática de metadatos para documentos digitales no estructurados, particularmente libros y textos digitalizados en formato PDF u otros similares. El proyecto busca identificar, de forma automática, información clave como título, autor(es), edición e ISBN, reduciendo significativamente el tiempo y esfuerzo necesarios para la catalogación, y permitiendo procesar un mayor volumen de materiales con recursos limitados.

La investigación combina dos enfoques complementarios: el infoque PLN clásico, basado en expresiones regulares, reconocimiento de patrones, algoritmos de clasificación, agrupamiento y detección de similitud, probado con éxito en libros digitales en español y distintos tipos de tipografía; y los modelos de lenguaje (LLM), mediante ingeniería de *prompts* y código en Python, aplicados con Qwen 2.5, adecuados para procesar documentos manuscritos, así como textos en lenguas indígenas o español antiguo. Este trabajo busca sentar las bases para un sistema robusto y escalable que contribuya a la preservación y acceso eficiente al patrimonio documental mexicano.

Autor(es): Belém Priego Sánchez y Ángel Hernández Esparza

Institución: UAM-Unidad Azcapotzalco

Título: Clasificación automática de tipos de despido en documentos laborales



Resumen: En México, la elevada incidencia de conflictos laborales, donde el 88% de las controversias están relacionadas con despidos injustificados, pone de manifiesto la necesidad de contar con un sistema que permita clasificar, de forma precisa y ágil, los documentos legales, facilitando así la toma de decisiones y el cumplimiento de la normativa

correspondiente. En esta investigación se analizan textos en materia laboral para identificar automáticamente diferentes tipos de despido o rescisiones de contrato en documentos legales laborales mediante técnicas de procesamiento de lenguaje natural y aprendizaje automático. Para dicho propósito, se desarrolló un sistema modular que asigna de manera automática las categorías de despido justificado, despido injustificado y despido colectivo. Además, cuenta con una interfaz web interactiva que permite la carga, visualización y consulta dinámica de resultados en tiempo real, facilitando la toma de decisiones. Esta propuesta permite mejorar la eficiencia de gestión documental y reducir significativamente los errores en los procesos de clasificación que habitualmente se usan en estas tareas, contribuyendo así a una toma de decisiones más precisas en entornos legales.

Autor(es): Regina Arahí Guerra Hernández

Institución: UNAM

Título: Hacia una ontología semántica de la música chilena: propuesta preliminar para la representación estructurada del conocimiento musical



Resumen: Este trabajo presenta una propuesta preliminar para el desarrollo de una ontología semántica destinada a la representación estructurada del conocimiento musical chileno contemporáneo. El objetivo es sistematizar, preservar y facilitar el acceso a información relevante sobre artistas, canciones, géneros, épocas, lugares, colaboraciones y temáticas presentes en la música chilena.

La ontología se construirá mediante el modelado conceptual en Protégé, identificando entidades clave y sus relaciones semánticas. La información se recopilará desde fuentes abiertas como catálogos discográficos, bases de datos culturales y letras de canciones, mediante técnicas de extracción como webscraping y uso de APIs. Estos datos serán normalizados y convertidos al formato RDF para su integración en un grafo de conocimiento, alojado en plataformas como GraphDB. El grafo podrá ser consultado mediante el lenguaje SPARQL, y se complementará con herramientas de visualización interactiva que permitan explorar el conocimiento de manera accesible e intuitiva.

La propuesta aborda desafíos metodológicos relacionados con la heterogeneidad de las fuentes y la especificidad cultural del vocabulario musical chileno. Finalmente, se plantea como una base para el desarrollo de futuras herramientas digitales que fomenten el análisis, la preservación y la difusión del patrimonio musical, así como su integración con iniciativas de datos en el ámbito hispanohablante.

Autor(es): Sergio Luis Ojeda Trueba y Gemma Bel Enguix

Institución: UNAM-IINGEN

Título: Clasificación de mitos por medio de métodos de *clustering*



Resumen: Este proyecto explora la recurrencia de mitemas —unidades narrativas compartidas en mitos de distintas culturas— mediante técnicas de Procesamiento de Lenguaje Natural (PLN). Se creó un corpus propio, el Primigenia Corpus, con 21 mitos de siete culturas (cristiana, maya, romana, hindú, nórdica, hopi y maorí) enfocados en tres mitemas: creación del mundo, creación de la humanidad y el diluvio. Los textos fueron preprocesados, traducidos al inglés y procesados mediante dos algoritmos de *clustering* distintitos: DBSCAN y K-Means, no sin antes llevar a cabo una vectorización TF-IDF y con BERT, respectivamente.

Los resultados muestran que los textos tienden a agruparse por cultura más que por mitemas, ya que es evidente una separación evidente entre los *clusters* de ambos métodos. Esto sugiere que la variabilidad léxica entre culturas y el bajo número de *tokens* por mitema afectaron el rendimiento del algoritmo. Se proponen mejoras como la expansión del corpus, otros métodos de vectorización y la normalización de entidades con herramientas de reconocimiento de nombres (NER).

Mesa 2. LENGUAS INDÍGENAS

Autor(es): Edén Espinoza Urzúa

Institución: UNAM

Título: Diseño e implementación de una aplicación de diccionarios parlantes para lenguas indígenas con integración de corpus y búsquedas inteligentes mediante el uso de inteligencia artificial



Resumen: En esta participación se aborda el desarrollo y el diseño e implementación de una aplicación autónoma de diccionarios parlantes para lenguas indígenas, integrando corpus lingüísticos multimodales, búsqueda inteligente y herramientas de inteligencia artificial. Ante la amenaza de extinción que enfrentan muchas lenguas originarias, el proyecto plantea una solución tecnológica que responda a necesidades lingüísticas específicas: morfología compleja, variación fonética y ausencia de recursos digitales adaptados.

A diferencia de plataformas existentes, esta propuesta busca superar limitaciones estructurales y funcionales, mediante el desarrollo de una arquitectura local, autónoma y escalable. El sistema incluirá segmentación morfológica automática, lematización, alineación audio-texto semi-automática y un motor de búsqueda fonético-semántico, todo ello implementado con tecnologías abiertas y accesibles.

La colaboración con el INAH garantiza la calidad lingüística del corpus, y se promueve una participación activa de las comunidades hablantes en la validación de contenidos. Además del impacto tecnológico y científico, el proyecto destaca por su dimensión social y educativa, al empoderar a las comunidades en la conservación de su patrimonio lingüístico y facilitar materiales interactivos para su enseñanza. En conjunto, la propuesta se inserta en la vanguardia de la lexicografía digital y la ingeniería del lenguaje con perspectiva intercultural.

Autor(es): Jean Carlos Buenfil Aguilar, Silvia Fernández Sabido, Jorge Reyes Magaña, Luis Basto Díaz, Luis Fernando Curi Quintal y Oscar Gerardo Sánchez Siordia

Institución: Universidad Autónoma de Yucatán

Título: Detectando el maya yucateco y el español con IA: una herramienta para responder en emergencias



Resumen: La gran diversidad lingüística en regiones indígenas puede generar situaciones críticas cuando se necesita atención médica o legal urgente y no se identifica el idioma que hablan las personas involucradas. Aunque existan intérpretes disponibles, ¿cómo saber a cuál contactar? Este trabajo aborda el problema al explorar la identificación automática entre el maya yucateco y el español mediante modelos de aprendizaje profundo.

Se usaron 3,820 audios: 2,116 en español y 1,704 en maya, divididos en fragmentos de 3 a 15 segundos y, cuando fue posible, balanceados por género. Se generaron espectrogramas RGB como entrada para los modelos. Se evaluaron tres arquitecturas: ResNet-152, ResNet-50 y MobileNetV2, comparando precisión, exactitud y eficiencia. ResNet-152 obtuvo el mejor desempeño (90.57% de exactitud y F1 de 0.9010). MobileNetV2, aunque menos precisa (87.17%), resultó más eficiente, ideal para entornos con recursos limitados.

Los resultados sugieren que se necesitan más datos, además de ampliar el estudio a otras lenguas con pocos recursos. El trabajo muestra el potencial del aprendizaje profundo para identificar idiomas indígenas y acelerar la búsqueda de intérpretes en emergencias.

Autor(es): María Fernanda Arzate Cárdenas y Antonio Reyes Pérez

Institución: Universidad Autónoma de Querétaro

Título: Análisis de sentimientos de indigenismos en el español de México



Resumen: En esta propuesta se presentan algunos avances de una investigación en activo relacionada con el análisis de sentimientos de indigenismos en el español mexicano, en particular, de préstamos, es decir, de unidades léxicas que han sido “tomadas” de otra lengua (Campbell, 1999, p. 58), en este caso, del náhuatl, el purépecha y el maya yucateco.

A partir de una búsqueda exploratoria en corpus, se identificaron varios contextos en los que la presencia de estos préstamos implica una carga valorativa. Dicha carga puede determinar que el contexto sea clasificado con polaridad positiva o negativa y, aunque el indigenismo no necesariamente es el ítem que determina tal polaridad, sí puede exponer un estigma social que contribuye a ella. En este sentido, se busca explorar qué polaridad adquieren los contextos en los que aparecen estos préstamos y analizar si la polaridad varía según el origen del préstamo.

La importancia del estudio radica en que podría revelar actitudes sociolingüísticas implícitas en los hablantes de español mexicano, así como posibles estigmas relacionados con las lenguas indígenas. Lo anterior contribuiría a visibilizar formas sutiles de discriminación lingüística presentes en los usos cotidianos del español asociados con las lenguas indígenas.

Mesa 3. EVALUACIÓN DE LLM

Autor(es): César Ricardo Segura Alejo, Gemma Belenguix y Helena Gómez Adorno

Institución: UNAM-IIMAS

Título: Análisis de la percepción en definiciones médicas en español y catalán mediante modelos basados en *transformers* y LLM



Resumen: La percepción que los niños tienen sobre los términos médicos puede reflejar aspectos emocionales, sociales y culturales, en especial al analizar los cambios ocurridos antes y después de eventos de gran impacto como la pandemia por covid-19. Este trabajo explora el uso de modelos de lenguaje a gran escala (LLM) y otros modelos basados en la arquitectura Transformer [1], para clasificar automáticamente la percepción (sentimiento) en definiciones de términos médicos escritas por estudiantes de nivel primaria de español y catalán con un enfoque *zero-shot*.

Se trabajó con un corpus compuesto por más de 30,000 definiciones médicas en español y catalán, recolectadas antes y después de la pandemia. Debido a que el corpus incluía definiciones en ambos idiomas y en diferentes periodos temporales, se dividió en subcorpus según la temporalidad (antes y después de la pandemia) y el idioma (español y catalán) para facilitar su análisis.

El objetivo fue identificar si existían diferencias significativas en la percepción (positiva, negativa o neutra) asociada a los términos, dependiendo del periodo de recolección. Para llevar a cabo el análisis, se utilizaron modelos basados en la arquitectura Transformer pre-entrenados, que incluyen tanto modelos de lenguaje a gran escala (LLM) como modelos más ligeros. Estos modelos están disponibles públicamente en plataformas como Hugging Face y Ollama [2]. En particular, se empleó el modelo *agentlans/multilingual-e5-small-alig-*

ned-sentiment[3], seleccionado por su entrenamiento específico en idioma catalán, según su documentación oficial.

Además, se evaluaron otras alternativas que también incluyen entrenamiento en catalán, como Abdel12/twitter-sentiment-analysis [4] y para corpus en español se utilizó un modelo entrenado únicamente para el idioma español nlptown/bert-base-multilingual-uncased-sentiment [5]. En cuanto a los modelos LLM, se realizó una evaluación mediante prueba y error con diferentes prompts, utilizando modelos como LLaMA3 y Mistral [6].

Una vez obtenido el etiquetado automático sobre los distintos corpus, se procedió a seleccionar muestras para ser etiquetar de forma manualmente, y las etiquetas automáticas, con el objetivo de calcular la matriz de confusión y evaluar métricas [7] de desempeño: recall, F1-score, accuracy, precision y specificity. Finalmente, se realizaron análisis de distribución de frecuencias por polaridad (positiva, neutra, negativa) y un análisis sobre la percepción asociada a términos específicos, lo cual permitió identificar patrones relevantes y cambios en la carga emocional.

En general, los modelos basados en Transformers muestran una capacidad aceptable para clasificar sentimientos. En particular, el modelo Abdel12/twitter-sentiment-analysis logró un desempeño razonable, a pesar de no estar entrenado específicamente con datos del dominio educativo o médico, por otro lado, el modelo entrenado exclusivamente en español alcanzó un desempeño excelente, con métricas sobresalientes en precisión, recall y F1 Score, lo que indica que la especialización en el idioma, en el caso de multilingual-e5-small-aligned-sentiment, si bien mostró fortalezas al identificar emociones negativas, tuvo dificultades al generalizar correctamente los casos positivos, esto sugiere que, aunque útil, su rendimiento es sensible a la carga semántica y polaridad del texto, finalmente los modelos LLM, mostraron un buen desempeño general, especialmente cuando se utilizaron prompts bien diseñados.

Los resultados muestran que el uso de modelos de lenguaje preentrenados, como los LLM, resulta viable para clasificar sentimientos en definiciones médicas escritas por niños, aunque su efectividad depende fuertemente del dominio y del ajuste previo. Los modelos específicos para el idioma español muestran mejor desempeño, mientras que los modelos multilingües pueden requerir ajustes complementarios para alcanzar resultados óptimos.

De esta forma, podemos identificar que el análisis resulta aceptable, a pesar de algunas terminologías ambiguas como confinamiento, que dependiendo de la definición podría catalogarse como positiva o negativa. Existen algunas definiciones donde es evidente que la percepción cambió, como en el caso de vacuna, que antes de la pandemia tenía una percepción negativa y pasó a ser positiva, o hospital, cuya percepción se tornó ligeramente más negativa posterior a la pandemia. Así, los LLM y modelos basados en *transformers* pue-

den ser de gran ayuda para analizar la percepción, incluso en textos breves y redactados por niños, sin requerir entrenamiento adicional.

**Autor(es): Judith Tavaréz Rodríguez, Fernando Sánchez Vega
y Adrián Pastor López Monroy**

Institución: Centro de Investigación en Matemáticas-Guanajuato

Título: Sesgos demográficos en la percepción de sexismo en modelos de lenguaje grandes



Resumen: El uso de modelos de lenguaje grandes (LLM) ha demostrado ser una herramienta que podría ayudar en la detección automática del sexismo. Estudios previos han demostrado que estos modelos contienen sesgos que no reflejan con precisión la realidad, especialmente para los grupos minoritarios. A pesar de los diversos esfuerzos para mitigarlos, esta tarea sigue siendo un desafío significativo debido a su naturaleza subjetiva.

En este trabajo, analizamos los sesgos demográficos presentes en algunos LLM, utilizando el conjunto de datos de tuits EXIST 2025. Este conjunto incluye anotaciones de seis anotadores con diferentes perfiles demográficos para cada tuit, lo que nos permite evaluar hasta qué punto los LLM pueden imitar las percepciones de estos grupos en la detección del sexismo. Además, realizamos un análisis estadístico para identificar qué características demográficas (edad, género) contribuyen con mayor eficacia a replicar la percepción del sexismo en un LLM.

Nuestros resultados muestran que, si bien los LLM pueden detectar el sexismo en cierta medida al considerar la opinión general de la población, no replican con precisión la diversidad de percepciones entre los diferentes grupos demográficos. Esto resalta la necesidad de modelos mejor calibrados que consideren la diversidad de perspectivas en diferentes poblaciones.

Autor(es): Alef Zaín Gama Sandoval

Institución: UNAM

Título: Qué tan lejos estamos de una inteligencia artificial generalizada



Resumen: Este documento es el resumen de un artículo de 14 cuartillas que trata sobre la distancia que aún nos separa de una inteligencia artificial general (AGI) a partir de un caso práctico: mejorar la precisión de grandes modelos de lenguaje (LLM) al alinear traducciones bíblicas con los textos hebreo y griego originales.

Primero se explica el trabajo especializado de la alineación —un proceso para brindar nuevos recursos lingüísticos en Software Bíblico Logos—, donde se describen cuatro tipos de

correspondencia ($1 \rightarrow 1$, $1 \rightarrow n$, $n \rightarrow 1$, $n \rightarrow n$) y la necesidad de etiquetar cada vínculo como primario, secundario, modismo o “no equivalente”.

Después se muestra una hipótesis de que los LLMs pueden realizar la alineación sin intervención humana. Se hizo un primer modelo de trabajo que mejoró hasta un tercer modelo que consiste en el uso automatizado de prompt + RAG (Retrieval-Augmented Generation) para alimentar la IA con datos de contexto (instrucciones, idiomas originales, alineaciones previas).

Posteriormente, se evaluaron los resultados de seis modelos (Claude 3.7, DeepSeek V3, Gemini 2.5, GPT-4o, Llama 3 y OpenAI o1) usando métricas de exhaustividad (todas las palabras de la traducción y de la fuente presentes), exactitud de WN_id (identificador de palabra fuente), corrección semántica y clasificación.

Los resultados muestran avances notables: Gemini 2.5 logró 100% de cobertura y exactitud, con tasa global de fidelidad del 100%; GPT-4o y o1 rondaron el 90-92%, mientras Llama 3 quedó muy rezagado. Sin embargo, se advierte sobre posibles sesgos en las pruebas modelo-a-modelo y la necesidad de conjuntos de validación independientes. También se señalan desafíos técnicos: consumo de memoria, tiempos de respuesta y complejidad para manejar excepciones idiomáticas.

En conclusión, sí hay un potencial de los LLM para tareas altamente especializadas que demuestran una posibilidad de una AGI, aunque no de forma autónoma, sino con la ayuda del ser humano, quien también debe aprender a “dialogar” con los modelos por medio de técnicas y herramientas que le permitan comprender mejor, de los LLM de las IA, su plano del contenido, o más bien, su plano del contenido *tokenizad*. Sin embargo, todavía falta integrar memoria de largo plazo, aprendizaje continuo, un continuo diálogo máquina-humano y muchos otros desafíos que nos acerquen, cada vez más, a dicha meta.

Autor(es): Ignacio Arroyo Fernández

Institución: Universidad Tecnológica de la Mixteca

Título: Evaluación de modelos estándar de atención para razonamiento semántico en bases de conocimiento biomédicas abiertas



Resumen: Este estudio examina si los modelos neuronales estándar de atención pueden realizar razonamiento en el nivel semántico a partir de bases de conocimiento de dominio abierto a un bajo costo. Se evaluaron arquitecturas con atención GRU, LSTM y Transformers standard utilizando tres bases de conocimiento y sus combinaciones: ConceptNet para sentido común, una base OpenIE genérica y otra extraída de 1,200 resúmenes PubMed sobre nueve enfermedades crónicas (52000 a 652000 tripletas de conocimiento en total).

Los modelos se entrenaron para predecir frases objeto a partir de pares sujeto-predicado. Además de las métricas tradicionales, se introdujo la Prueba de Preferencia Seleccional Basada en Significado (MSPT), la cual enmascara frases objeto y mide ganancias de similitud coseno contra referencias reales y aleatorias. Un análisis con bosques aleatorios mostró una influencia importante del número de capas y dropout en la pérdida de predicción. Además, un análisis cualitativo de los mapas de atención y la anotación manual de marcos semánticos, roles y preferencias de selección mostró los patrones de errores semánticos de los modelos, incluyendo la sobregeneralización sistemática en roles y preferencias.

Los resultados muestran que la mejor red GRU logró una brecha MSPT del 20.7%, mientras que el mejor *transformer* alcanzó 9.9%, ambos muy significativos, superando las líneas de base LSTM y demostrando la viabilidad del razonamiento semántico biomédico con bases de conocimiento fáciles de obtener (aunque ruidosas) y arquitecturas neuronales ligeras.

Mesa 4. CORPUS LINGÜÍSTICOS

Autor(es): Pamela Ábrego Pérez

Institución: El Colegio de México

Título: Experiencia en la construcción del corpus “Los debates electorales en México”: desafíos y aprendizajes desde la práctica lingüística



Resumen: Este trabajo presenta una reflexión desde la experiencia individual como colaboradora en el proyecto de lingüística de corpus “Los debates electorales en México”, dirigido por la Dra. María Eugenia Vázquez Laslop en El Colegio de México. El proyecto reúne a un equipo multidisciplinario encargado de recopilar, transcribir y etiquetar debates electorales en México desde 1994.

Mi participación se ha centrado en la localización de materiales audiovisuales y la elaboración de transcripciones anotadas en formato XML, lo que permite que el corpus sea procesado para distintos fines: consulta general, análisis lingüísticos, políticos, sociológicos, entre otros. No todas las etiquetas categorizan elementos lingüísticos, sino apenas unas cuantas. Desde esta labor, abordaré las principales dificultades encontradas en el proceso: el acceso a los videos, la baja calidad de algunos archivos y la necesidad de establecer criterios claros y compartidos para la anotación manual, a fin de mantener la coherencia del corpus.

Aunque las soluciones técnicas y de codificación han sido desarrolladas por un equipo especializado, esta presentación se enfoca en los procesos prácticos del trabajo lingüístico, el diálogo constante entre participantes y la búsqueda de consensos para la estructura del corpus.

Autor(es): Carlos Alberto Rivera Flores

Institución: UNAM-FFyL

Título: El desafío del etiquetado del corpus de cómics mexicanos



Resumen: Uno de los principales problemas al crear el corpus de cómics mexicanos es su etiquetado. Nuestro punto de partida fue el Comic Book Markup Language (CBML), creado por la Universidad de Indiana, que posee muchísimas etiquetas tanto para las partes del cómic como para los metadatos. Decidimos adoptar algunas etiquetas del CBML, pero adecuarlas y crear unas propias que abarquen las necesidades que tenemos, siguiendo las bases de XML. El etiquetado está sujeto a cambios, pues somos conscientes de que será necesario replantear o crear etiquetas.

El etiquetado está dividido en dos partes: los metadatos y el cuerpo del cómic. La primera sección contiene información bibliográfica propia de cada título extraída de la base de datos Pepines. Respecto a la segunda, contiene la mayor cantidad de etiquetas y variables, debido a los múltiples matices por abarcar. Por ejemplo, se señala quién es el personaje enunciador del diálogo, si se trata de un pensamiento, un sueño, un grito o un murmullo. Para extraer el texto empleamos el OCR de Qwen, una IA que agiliza la extracción del texto y su etiquetado. Parte de su labor es entregar el cómic en un archivo XML para integrarlo al corpus.

**Autor(es): Ximena Gutierrez Medina, Fernando Sánchez Vega
y Adrián Pastor López Monroy**

Institución: Instituto Tecnológico Superior de Guanajuato

Título: Cuando la emoción guía la voz ciudadana: análisis de preguntas al gobierno



Resumen: En este trabajo se examina un conjunto de preguntas formuladas por la ciudadanía a diversas instancias gubernamentales. A partir de este corpus, se identificaron los errores más frecuentes en la construcción y formalización de dichas preguntas, tales como la falta de claridad, el uso inadecuado del lenguaje o la ausencia de datos contextuales esenciales.

Asimismo, se analizó el rol de las emociones y sentimientos expresados por los ciudadanos al interpelar al gobierno, observándose una fuerte presencia de frustración e indignación, lo que puede condicionar la estructura y el tono del mensaje. El estudio también clasificó las principales temáticas abordadas, destacando asuntos como la seguridad, la salud pública y la corrupción.

En conjunto, los hallazgos revelan una participación ciudadana activa, aunque cargada de ambigüedades y emociones, lo cual puede influir negativamente en la efectividad comuni-

cativa y en la disposición del Estado a responder, abriendo un campo de reflexión sobre los canales de diálogo institucional.

Mesa 5. ANÁLISIS TEXTUAL

Autor(es): César Alejandro Cabrera Peralta

Institución: UNAM-FFyL

Título: Distributional Semantics Toolkit: plataforma integrada para el análisis y visualización de textos mediante modelos distribucionales



Resumen: Distributional Semantics Toolkit es una biblioteca de código abierto escrita en Python para el análisis textual mediante métodos distribucionales. Está basada en el libro *Distributional Semantics*, de Alessandro Lenci y Magnus Sahlgreen, y su objetivo es gradualmente incorporar los algoritmos ahí descritos, comúnmente usados en la semántica distribucional.

Actualmente, incluye clases para el preprocesamiento de texto (tokenización, lematización, extracción de vocabulario, etiquetado POS), la extracción de colocaciones (dirigidas y no dirigidas), generación de matrices (matriz de coocurrencias y matriz documento-término) y su ponderación (PMI y PPMI).

También permite la generación de word embeddings mediante modelos de reducción de dimensionalidad y de predicción (Descomposición de Valores Singulares, Análisis de Componentes Principales, Word2Vec, FastText), el cálculo de distancias geométricas (euclídea, Manhattan, similitud del coseno, términos relacionados) y visualización de *embeddings* (método del codo, coeficiente de Silhouette, gráficos en dos y tres dimensiones).

Está diseñada para ser accesible a lingüistas con conocimientos limitados de Python, y sus métodos pueden emplearse de manera independiente o integrarse en *pipelines*, de forma similar a spaCy, por medio de una interfaz sencilla, integrada y flexible. La herramienta busca facilitar investigaciones en semántica computacional, humanidades digitales, análisis del discurso, aprendizaje automático y lingüística aplicada, especialmente en contextos educativos.

Autor(es): Kapioma Villarreal Haro, Fernando Sánchez Vega y Adrián Pastor López Monroy

Institución: Centro de Investigación en Matemáticas

Título: Contexto para guiar, razonamiento para aprender: una aproximación híbrida de *fine-tuning* con explicaciones de LLM para detección de sexismo



Resumen: La detección automática de sexismo en textos de redes sociales es una tarea altamente sensible que se enfrenta a retos como la dependencia del contexto o lenguaje ruidoso. Los modelos compactos basados en transformers como BERT o XML-RoBERTa, aunque populares para clasificación en tareas como esta, carecen de capacidades de contextualización profunda.

Por otro lado, a pesar de la potencia los Modelos de Lenguaje Generativos de gran escala y sus habilidades como el “razonamiento”, siguen sin ser suficientemente precisos para estas tareas en escenarios *few-shot* o *zero-shot*. Proponemos un esquema híbrido que aprovecha lo mejor de ambos tipos de modelos para la detección de sexismo en redes sociales.

Nuestra metodología consta de tres etapas: uso de *prompts* contextualizados que incorporan información relevante y alineada con la tarea antes de la clasificación, generación de textos de justificación para mostrar el razonamiento del modelo detrás de la clasificación, y uso de estas justificaciones como expansión de conocimiento para el entrenamiento de modelos de clasificación compactos. En este esquema, modelos ligeros integran capacidades de razonamiento de LLM grandes que han sido contextualizados adecuadamente, mejorando tanto *baselines* generativos simples como esquemas de *fine-tuning* tradicionales.

Autor(es): Iván Díaz, Jessica Barco, Joana Hernández, Edgar Lee Romero y Gemma Bel-Enguix

Institución: UNAM-IINGEN

Título: Detección de subjetividad mediante modelos simples de clasificación BERT



Resumen: El presente trabajo es resultado de la participación del equipo cepanca_UNAM en la competencia internacional CheckThat! 2025, específicamente en la tarea de detección de subjetividad en artículos de noticias. CheckThat es uno de los laboratorios del CLEF (Conference and Labs of the Evaluation Forum). El objetivo de la tarea del concurso fue distinguir automáticamente si una oración expresa una opinión subjetiva u objetiva, por medio de un sistema de clasificación binaria.

Se realizó un análisis cuantitativo para conocer el estado de los datos proporcionados con los que se entrenaron los modelos basados en *transformer*. El modelo preentrenado seleccionado fue BERT aplicado en tres idiomas: alemán, italiano e inglés, donde se comparó su

desempeño en contraste con el modelo clásico de regresión logística; además se adaptaron los modelos con análisis de etiquetas sintácticas y atención al desequilibrio de clases (obj / subj), ajustando el tamaño de lote y la longitud máxima de entrada.

Autor(es): Diego Hernández Bustamante y Helena Gómez Adorno

Institución: UNAM-FES Aragón

Título: Uso de modelos de lenguaje generativos para la clasificación de entidades médicas en casos clínicos en español



Resumen: La identificación automática de entidades médicas en textos clínicos es una tarea clave para mejorar el acceso, la organización y el análisis de la información contenida en historias clínicas electrónicas. Este tipo de extracción permite facilitar tareas como la codificación médica, el apoyo al diagnóstico y la investigación clínica [1] [2]. En este trabajo se evalúa el desempeño de modelos de lenguaje de gran escala (LLMs) en la identificación de entidades médicas en español.

Los modelos fueron evaluados en un escenario *zero-shot*, es decir, sin entrenamiento adicional, con el objetivo de analizar su capacidad de generalización en tareas de reconocimiento de entidades médicas. Para ello, se trabajó con tres corpus de acceso libre escritos en español: SympTEMIST [3], DisTEMIST [4] y MedProcNer [5], los cuales utilizan los mismos reportes de caso, pero etiquetan entidades distintas síntomas, enfermedades y procedimientos, respectivamente. Estos corpus fueron unificados mediante la integración de sus anotaciones en un único conjunto de datos, preservando la segmentación por tipo de entidad.

Se diseñaron tres *prompts* distintos para extraer las entidades médicas de los reportes de caso y devolver dichas entidades en forma de listas separadas (una por tipo de entidad). Se utilizaron los modelos de: Gemini, LLama 3.1 y 3.2, Qwen 2 y 2.5, Internlm 2.5y 3 y Mistral. Como este corpus era de acceso libre, se pudo experimentar con modelos privativos (Gemini); pero para tener un mayor control de la privacidad, es importante usar modelos *opensource*. Una vez generadas las salidas, éstas se limpiaron y normalizaron para después evaluar su rendimiento mediante las métricas de precisión, recall y F1-score. Finalmente, se comparó el desempeño de los distintos LLM analizados para identificar cuál obtuvo los mejores resultados.

Los resultados indican que, pese a sus capacidades multifuncionales, los modelos de lenguaje a gran escala (LLM) aún no constituyen una solución efectiva para el Reconocimiento de Entidades Nombradas (NER), mostrando un rendimiento significativamente inferior en comparación con modelos especializados (fine-tuned models) en dominios clínicos.

Mesa 6. PRECISIÓN LINGÜÍSTICA DE LLM

Autor(es): Abigail Vargas Hernandez, Fernando Sánchez Vega
y Adrián Pastor López Monroy

Institución: Instituto Tecnológico Superior de Guanajuato

Título: Cómo conjugan el español los grandes modelos del lenguaje



Resumen: El trabajo analiza la capacidad de los grandes modelos del lenguaje (LLM) para conjugar verbos en español, con el objetivo de evaluar su competencia morfológica en tareas lingüísticas específicas. Se exploró la conjugación de una amplia gama de verbos, incluyendo tanto formas regulares como irregulares, y se diseñaron pruebas para observar cómo los modelos enfrentan distintos desafíos morfológicos.

Los resultados fueron categorizados según una tipología verbal, lo que permitió identificar patrones consistentes de aciertos y errores. Se emplearon diversas estrategias de evaluación: generación sin ejemplos previos (*zero-shot*), con pocos ejemplos (*few-shot*) y mediante la inyección explícita de conocimiento lingüístico. Estas estrategias se compararon para medir su efectividad y robustez ante distintos tipos de verbos y configuraciones gramaticales. La evaluación reveló variaciones significativas en el rendimiento de los modelos según el enfoque utilizado y el tipo de verbo evaluado.

Finalmente, el estudio presenta un conjunto de recomendaciones prácticas para mejorar el desempeño de los LLM en tareas de conjugación verbal en español, destacando la utilidad de técnicas de ingeniería de *prompts* y la importancia de un entrenamiento más explícito en fenómenos morfológicos complejos.

Autor(es): Alejandra Mitzi Castellón Flores, Armando Quetzalcóatl Angulo Chavira,
Alejandra Ciria y Natalia Arias Trejo

Institución: UNAM-Facultad de Psicología

Título: Benchmark oracional para la predicción del lenguaje en español mexicano



Resumen: La anticipación léxica —la predicción de la siguiente palabra en un contexto— constituye un mecanismo fundamental tanto en modelos probabilísticos de lenguaje como en el procesamiento cognitivo humano (Dell & Chang, 2014). Para evaluar este fenómeno se desarrolló un corpus de normas de finalización de oraciones en español de México, formando un recurso con 2,925 contextos oracionales.

Se recabaron datos de 1,470 participantes adultos, quienes generaron continuaciones libres para cada oración. A partir de estas respuestas, se calculó la probabilidad de cierre,

y además medidas informacionales como surprisal ($-\log p$) y entropía, capturando la incertidumbre del contexto. Asimismo, se anotaron propiedades morfosintácticas y semánticas de cada contexto oracional (lema, categoría, género, número, función, concreción, frecuencia, longitud).

Las oraciones se clasificaron según predictibilidad (alta, media, baja) y nivel de incertidumbre, brindando una rica variabilidad experimental. Este recurso permite evaluar y comparar modelos autoregresivos, métricas de sorpresa léxica y sistemas de autocompletado entrenados en corpus de gran escala. Adicionalmente, facilita la selección de estímulos controlados para estudios con seguimiento ocular, EEG o lectura automatizada.

Al reflejar rasgos dialectales del español mexicano y registrar múltiples métricas predictivas, complementa recursos peninsulares y panhispánicos, habilitando análisis transculturales en lingüística computacional.

Mesa 7. ANÁLISIS PSICOLINGÜÍSTICO

Autor(es): Armando Quetzalcóatl Angulo Chavira, Alejandra Mitzi Castellón Flores, Vladimir Huerta-Chavez y Luis Ángel Llamas Alonso

Institución: UNAM-Facultad de psicología

Título: ANEW-MX: adaptación mexicana de las normas afectivas para palabras



Resumen: Presentamos la primera adaptación mexicana del repertorio Affective Norms for English Words (ANEW), que ofrece normas afectivas para 1028 palabras de uso cotidiano. Un número de 753 hablantes mexicanos (edad media 25 años) calificaron cada término con la escala pictórica Self-Assessment Manikin en línea. Para cada palabra estimamos tres dimensiones emocionales: valencia, que indica cuán agradable o desagradable resulta leerla; activación, la energía fisiológica que provoca, desde la calma hasta la excitación; y dominancia, la sensación de control que acompaña a la emoción, comparable a sentirse empoderado o abrumado. Estas métricas permiten realizar una descripción multidimensional de matices que la simple etiqueta “positiva” o “negativa” no captura.

Las normas muestran alta consistencia interna y reproducen relaciones no lineales entre dimensiones, como la curva en U entre valencia y activación, clave para modelar dinámicas emocionales. Además, se observó una relación inversa entre activación y dominancia, útil para inferir polaridad afectiva más allá de clasificaciones binarias. Esta adaptación emplea vocabulario con alta frecuencia y relevancia en México, e incluye métricas léxicas como frecuencia, longitud y categoría gramatical. Estas propiedades permiten integrar el recurso como entrada o supervisión en modelos computacionales para análisis de sentimiento, clasificación emocional, generación de lenguaje afectivo o ajuste de *embeddings* contextuales.

**Autor(es): Guillermo Segura Gómez, Adrián Pastor López Monroy
Fernando Sánchez-Vega y Alejandro Rosales Pérez**

Institución: Centro de Investigación en Matemáticas

Título: Optimización evolutiva de *prompts* para grandes modelos de lenguaje aplicado en tareas de detección emocional



Resumen: El diseño adecuado de *prompts* influye directamente en el rendimiento de los modelos de lenguaje (LLM), especialmente en tareas complejas como la detección de emociones. En este trabajo, proponemos un enfoque evolutivo para optimizar *prompts* mediante algoritmos genéticos, incorporando un esquema de mutación sistemática. A diferencia de las mutaciones aleatorias tradicionales, aquí se introducen transformaciones lingüísticas controladas, como la reformulación semántica, la especificación de contexto y la simplificación, con la idea de mejorar paso a paso la calidad de los *prompts* generados. Aplicamos esta metodología en el contexto del SemEval 2025 Task 11A, una competencia enfocada en identificar la emoción experimentada por el autor de un texto, más allá de la emoción que podría percibir un lector.

Nuestros experimentos muestran que las mutaciones sistemáticas permiten generar prompts más eficaces y estables, mejorando métricas como F1-score macro y micro, sobre todo en emociones complejas como “miedo” y “alegría”. Además, el análisis de supervivencia de mutaciones sugiere que algunas transformaciones estructuradas, en particular, la especificación de contexto, tienen mayor probabilidad de derivar en prompts exitosos. Este enfoque muestra que la evolución dirigida de prompts no sólo compite con métodos tradicionales, sino que también abre una vía escalable, reproducible y automatizada para el diseño de instrucciones en modelos de lenguaje.

Autor(es): Roberto Huerta Ponce, Helena Gómez Adorno y Gemma Bel Enguix

Institución: UNAM-IIMAS

Título: Exploración de la efectividad de modelos ligeros combinados para la detección de señales de salud mental en redes sociales



Resumen: Este trabajo presenta un enfoque eficiente para la clasificación automática de textos en redes sociales según categorías de riesgo relacionadas con la salud mental, tema que ha cobrado relevancia en los últimos años por el incremento de expresiones emocionales en línea. Se parte del uso combinado de representaciones lingüísticas complementarias: análisis de frecuencia de términos, análisis de sentimientos y modelado de tópicos. A partir de estas representaciones, se construye un ensamble de modelos clásicos de aprendizaje automático (Random Forest, MLP, KNN, entre otros), optimizados individualmente mediante búsqueda de hiperparámetros con el uso de técnicas especializadas. Las predic-

ciones generadas por estos modelos son integradas mediante un mecanismo de ensamble ponderado, afinado mediante un proceso de ablación iterativa tipo *greedy*, con el objetivo de reducir la complejidad sin comprometer el rendimiento. Adicionalmente, se emplea la herramienta LIME para analizar la contribución individual de cada modelo dentro del ensamble, permitiendo interpretar cómo cada representación aporta distintas señales al proceso de clasificación. Finalmente, se compara el desempeño de la propuesta con un modelo basado en Transformers (DistilRoBERTa), utilizando validación cruzada y métricas de eficiencia computacional. La arquitectura propuesta retoma el mecanismo de atención de características de trabajos previos, adaptándolo a las representaciones lingüísticas utilizadas en este trabajo (frecuencia de términos, análisis de sentimientos y modelado de tópicos), con el fin de ponderar la contribución de cada modelo especializado en dichas representaciones.

Para el entrenamiento y evaluación se utilizó un conjunto de datos de libre acceso, el cual contiene textos etiquetados en categorías relacionadas con la salud mental. Este mismo conjunto fue empleado para entrenar y comparar el desempeño de la arquitectura propuesta frente al modelo basado en DistilRoBERTa.

**Autor(es): Francisco Manzo Méndez, Fernando Sánchez Vega
y Adrián Pastor López Monroy**

Institución: Instituto Tecnológico Superior de Guanajuato

Título: Factores textuales relevantes para mejorar la detección de la depresión



Resumen: En este trabajo se analizan las capacidades de dos tipos de modelos de lenguaje: los autoregresivos generativos y los de clasificación bidireccionales, para identificar la depresión en conversaciones. Se evalúa el desempeño de ambos modelos en tareas de identificación automática de indicios de depresión a partir de textos conversacionales, comparando su precisión, sensibilidad y capacidad de generalización. Asimismo, se propone una metodología compuesta que combina las fortalezas de ambos enfoques: la habilidad generativa de los modelos autoregresivos y la precisión contextual de los modelos bidireccionales.

Esta estrategia integrada permite una mejor identificación de patrones lingüísticos asociados con estados depresivos, lo que a su vez posibilita la detección de factores textuales clave relacionados con la depresión. Entre estos destacan el uso de expresiones autorreferenciales negativas, patrones de aislamiento y marcadores de desesperanza. Dichos factores pueden ser detectados automáticamente y constituyen una base sólida para futuras mejoras en sistemas de apoyo a la detección temprana de trastornos depresivos.

Mesa 8. LINGÜÍSTICA FORENSE

Autor(es): Vanessa Díaz Rodríguez y Antonio Reyes Pérez

Institución: Universidad Autónoma de Querétaro

Título: Would Groomers Know Better “Would”? A Linguistic Analysis of Online Grooming



Resumen: El *grooming online* es un acto mediante el cual un adulto establece contacto con un menor con fines sexuales, generalmente a través de plataformas digitales (Lorenzo Dus e Izura, 2017). Este trabajo se centra en el estudio de un elemento discursivo recurrente en interacciones clasificadas como grooming: el verbo modal *would*.

Para ello, se usó el *dataset* PAN12-Sexual Predator Identification (Inches & Crestani, 2012), el cual se ha constituido como un referente para tareas relacionadas con la identificación de predadores sexuales. Mediante una metodología guiada por corpus, se han identificado comportamientos de esta forma no reportados en la literatura.

Entre ellos, se ha observado que el verbo modal *would* presenta diversas funciones en interacciones de índole sexual, tales como marcador de condicionalidad, forma de subjuntivo, mitigación, entre otras. Adicionalmente, se han encontrado secuencias muy recurrentes, tales como *would + u*, *would + love*, *would + like*, así como su forma negativa e, incluso, su forma intencionalmente mal escrita *wud*, lo que revela que *would* puede ser usado como marcador de este tipo de interacciones.

Se proyecta que los hallazgos contribuyan directamente al campo del procesamiento de lenguaje natural y de la lingüística forense para apoyar al desarrollo de herramientas orientadas a la prevención del delito, el acceso a una vida libre de violencia, entre otros.

Autor(es): Ricardo Ulises Fuentes Yáñez, Joana Estefanía Hernández Rebollo, Carlos Alberto Rivera Flores y Fernanda Denisse García Cano Muñoz

Institución: UNAM

Título: Una aproximación exploratoria en las diferencias entre la voz generada por IA y la voz natural



Resumen: El uso de inteligencia artificial (IA) con el fin de imitar y generar voces digitalmente comienza a ser cada vez más frecuente en la actualidad, lo que produce ciertas cuestiones al realizar un análisis forense sobre voces dubitadas. En esta investigación exploratoria buscamos identificar las diferencias entre los parámetros acústicos de vocales emitidas por una IA traductora y una persona real perceptibles en un análisis realizado con el *software* Praat.

Dicha investigación se hizo con el fin de determinar si la emisión de vocales es un factor con el que se pueda diferenciar ambas voces, por lo que nos dimos a la tarea de obtener la frecuencia fundamental y los formantes vocálicos de ambas voces, lo que arrojó como resultados una menor duración de estos sonidos en el análisis a la voz artificial en comparación con la voz humana y otros datos de interés. La presente investigación exploratoria tiene el objetivo principal de señalar las diferencias en los parámetros acústicos de dos grabaciones de voz distintas, una perteneciente a una voz natural y otra producida con una IA que imite a la voz natural.

Autor(es): Fernanda López Escobedo

Institución: UNAM-Escuela Nacional de Ciencias Forenses

Título: Aplicaciones forenses de la lingüística computacional: del análisis de voz a la detección de plagio en la era de la inteligencia artificial



Resumen: El uso de herramientas computacionales ha impulsado significativamente el desarrollo científico en diversas áreas, y la lingüística forense no ha sido la excepción. Desde mediados del siglo XX, cuando el análisis de medidas estilométricas se basaba en modelos univariados, esta disciplina ha evolucionado hasta incorporar el uso de redes neuronales profundas para identificar al autor de un texto o una grabación.

Si bien el desarrollo de la lingüística computacional ha tenido un impacto positivo en el ámbito forense, también plantea retos importantes con el reciente avance de la inteligencia artificial. Entre ellos se encuentran, por ejemplo, la clonación de voces, que ha dado lugar a grandes fraudes, y la generación de sentencias basadas en normas inexistentes generadas por sistemas de inteligencia artificial.

El objetivo de esta ponencia es plantear los retos que enfrenta el campo de la lingüística forense en la era de la inteligencia artificial y cómo estos mismos desafíos pueden convertirse en aliados para la resolución de casos forenses en los que se involucre una evidencia oral o textual.

Autor(es): Xounely Chandel Dávila Gutiérrez

**Institución: Organización de Estados Iberoamericanos
para la Educación, la Ciencia y la Cultura**

Título: Situación epistémica del multilingüismo y su enfoque estructural ante el racismo lingüístico



Resumen: Las comunidades y pueblos indígenas, campesinas y afrodescendientes de México han construido espacios de encuentro y de vinculación para compartir experiencias,

luchas e historias a lo largo de la vida. En estas narrativas se comparten los sentires, los vínculos que existen con los territorios, y sobre todo visibilizan las violencias estructurales y coloniales que ha perdurado en las comunidades y que se representan en las luchas y los espacios de resistencia.

Es de vital importancia situar las problemáticas a las que se enfrentan las lenguas e idiomas indígenas las cuales se ven atravesadas por sistemas de opresión como el racismo, clasismo, la subordinación lingüística, y la discriminación.

La situación epistémica sobre el multilingüismo requiere contextualizarse trayendo a la discusión que la práctica de una o más lenguas en un espacio no debe influir en el desplazamiento de una lengua por otra, sino, reconocer los niveles de multilingüismo por medio de la situación contextual de las comunidades.

Asimismo, es importante reconocer el papel sustancial que tiene la educación comunitaria y popular para el acompañamiento de procesos pedagógicos que busquen visibilizar las deudas históricas que existen con respecto a reconocer otros procesos de construir el conocimiento y la vida misma.

Autor(es): Ismael Gómez García

**Institución: Organización de Estados Iberoamericanos para la Educación,
la Ciencia y la Cultura**

**Título: Superinteligencia y Superalianzas: Ética de la IA y el
futuro de la cooperación internacional**



Resumen: En 2019 BERT sorprendió al mundo del procesamiento de lenguaje natural, pero el giro copernicano en la inteligencia artificial llegó en noviembre de 2022 con el lanzamiento de ChatGPT 3.5, que desató una fiebre del oro y situó la IA de frontera exclusivamente en manos de grandes empresas con capitalizaciones millonarias, cuya inversión, infraestructura y volumen de datos marginan a poblaciones enteras y lenguas minoritarias.

Se plantea si la ética de la IA es un prólogo o un epílogo, pues mientras la narrativa dominante se centra en los sesgos y en la representatividad de minorías, quedan fuera enfoques como la ética empresarial, medioambiental o decolonial. Surgen propuestas como la superinteligencia anunciada por Meta en junio, donde Mark Zuckerberg afirma que acelerará el progreso humano, pero omite referencias a la educación y al cambio climático, mostrando que la superinteligencia se concibe como un producto dirigido a clientes individuales, sustituyendo los marcos colectivos de la ética por la satisfacción personal.

Así, el fetichismo de la superinteligencia la presenta como la solución a los grandes problemas de la humanidad cuando lo que falta no son ideas, sino voluntad global para forjar alianzas. Por ello, organizaciones como la OEI proponen superalianzas como alternativa para desarrollar éticas globales y cooperativas, en lugar de depender de soluciones privadas orientadas al consumo.



Agradecimientos

Instituto de Investigaciones Bibliográficas de la UNAM

Organización de Estados Iberoamericanos para la Educación,
la Ciencia y la Cultura (OEI)

Proyecto PAPIIT “Extracción automatizada de información para la
catalogación del patrimonio documental” IG400725

Proyecto PAPIIT “Creación de un acervo digital náhuatl en la
biblioteca nacional” IG400325

Proyecto “SECIHTI IH-2025-I-106: Calibración y validación de la herramienta interactiva para
comparación forense de voz ForVoz”

Proyecto CONAHCYT “Grafos conceptuales para la construcción de
diccionarios inversos en áreas de especialidad” CF-2023-G-64.

AGRADECIMIENTOS

Dr. Leonardo Lomelí Vanegas
Rector

Dra. Patricia Dolores Dávila Aranda
Secretaria General

Mtro. Hugo Concha Cantú
Abogado General

Mtro. Tomás Humberto Rubio Pérez
Secretario Administrativo

Dra. Diana Tamara Martínez Ruiz
Secretaría de Desarrollo Institucional

Lic. Raúl Arcenio Aguilar Tamayo
Secretario de Prevención, Atención y Seguridad Universitaria

Dr. María Soledad Funes Argüello
Coordinadora de la Investigación Científica

Dr. Miguel Armando López Leyva
Cordinador de Humanidades

Dr. Norma Blazquez Graf
Cordinadora para la Igualdad de Género

Dra. Rosa Beltrán Álvarez
Cordinadodra de Difusión Cultural

Mtro. Nestor Martínez Cristo
Director General de Comunicación Social

Mtro. Rodolfo González Fernández
Director de Información

COMITÉ ORGANIZADOR



Gerardo Eugenio Sierra Martínez
Grupo Ingeniería Lingüística
Instituto de Ingeniería UNAM

Fernando López Escobedo
Escuela Nacional de Ciencias Forenses

Agradecimiento
A Conahcyt,
Proyecto CF-2023-G-64