

Thot: Sistema de Sintetización de Información en la Sala Mexica del Museo Nacional de Antropología mediante Técnicas de Lenguaje Natural

Ingeniería de Software Para Sistemas Intelgentes

Diego Castillo Reyes
Marthon Leobardo Yáñez Martínez
Aldo Escamilla Resendiz

17 de septiembre de 2024

Índice general

1.	Introducción	2
2.	Problemática	2
3.	Objetivos	2
4.	Solución de Propuesta	3
5.	Alcance del Proyecto	3
6.	Análisis de Requerimientos	5
7.	Calendarización de Actividades	6
8.	Cálculo del Proyecto	8
8.1.	Métricas de Desempeño del Proyecto Thot	8
8.2.	Métricas de Calidad del Proyecto Thot	8
8.3.	Métricas de Costos del Proyecto Thot	9
8.4.	Métricas de Productividad del Proyecto Thot	9
8.5.	Métricas de Riesgo del Proyecto Thot	9

1. Introducción

El Museo Nacional de Antropología es uno de los recintos culturales más importantes de México y el mundo, albergando una vasta colección de objetos históricos que representan la rica herencia cultural de las civilizaciones precolombinas. En particular, la sala Mexica atrae a miles de visitantes interesados en aprender sobre esta civilización clave en la historia de Mesoamérica. Sin embargo, la cantidad de información disponible puede resultar abrumadora para muchos, lo que dificulta una comprensión accesible y rápida de los datos más relevantes.

En este contexto, surge el proyecto Thot, un Sistema de Sintetización de Información que tiene como objetivo ofrecer resúmenes concisos y precisos de los contenidos presentados en la sala Mexica del Museo Nacional de Antropología. A través del uso de técnicas avanzadas de Procesamiento de Lenguaje Natural (NLP), Thot permite extraer, sintetizar y presentar la información más importante, facilitando una experiencia educativa más accesible y personalizada para los visitantes.

Además, Thot incluye una actividad interactiva que busca enriquecer la experiencia del usuario. Al finalizar la síntesis de la información, el sistema presenta una trivia relacionada con los temas abordados, diseñada para que los visitantes puedan participar en un juego entretenido y educativo entre amigos y familiares.

Este sistema no solo busca enriquecer la visita al museo, sino también explorar nuevas formas de aplicar la inteligencia artificial en la preservación y divulgación del patrimonio cultural. Thot se posiciona como una herramienta innovadora para museos, donde la cantidad de información puede convertirse en una barrera para el entendimiento profundo de las colecciones, y al mismo tiempo fomenta la interacción y la diversión educativa.

2. Problemática

En la actualidad, los museos juegan un papel esencial en la educación y la preservación del patrimonio cultural. Sin embargo, uno de los principales desafíos que enfrentan es la sobrecarga de información que puede afectar la experiencia de los visitantes. La sala Mexica del Museo Nacional de Antropología es un ejemplo claro, donde la vasta cantidad de objetos y datos históricos puede resultar abrumadora, dificultando la asimilación rápida y efectiva de los contenidos expuestos.

Este problema se agrava cuando los visitantes buscan una experiencia más profunda o personalizada y no cuentan con el tiempo necesario para analizar toda la información. Las visitas guiadas pueden proporcionar un contexto valioso, pero están limitadas por horarios y la capacidad de personal disponible. Además, la información escrita en carteles o paneles a menudo carece de un formato que se ajuste a las necesidades individuales de los visitantes.

Adicionalmente, los museos enfrentan el desafío de atraer y retener el interés de un público más joven, acostumbrado a experiencias interactivas y dinámicas. La mayoría de las exhibiciones actuales son unidireccionales, lo que limita la interacción y participación activa de los visitantes, particularmente en grupos familiares o de amigos que buscan compartir una experiencia colectiva más entretenida y competitiva.

Por tanto, surge la necesidad de desarrollar herramientas que, mediante el uso de tecnologías innovadoras, ofrezcan soluciones que no solo faciliten el acceso a la información clave, sino que también proporcionen una experiencia interactiva y personalizada que motive a los visitantes a involucrarse activamente en su aprendizaje y a compartir ese proceso con otros.

3. Objetivos

1. Facilitar la comprensión del contenido histórico y cultural de la sala Mexica del Museo Nacional de Antropología mediante técnicas avanzadas de procesamiento de lenguaje natural.
2. Promover la interacción educativa y cultural entre los visitantes del museo, integrando actividades dinámicas como trivias y juegos que fomenten la participación activa.
3. Aplicar inteligencia artificial en la síntesis de información, creando experiencias más atractivas y personalizadas para los usuarios.

4. Solución de Propuesta

Thot es un Sistema de Sintetización de Información que utiliza técnicas avanzadas de Procesamiento de Lenguaje Natural (NLP) para extraer, resumir y presentar los contenidos de la sala Mexica del Museo Nacional de Antropología de manera accesible y personalizada. A través de una interfaz interactiva, los visitantes pueden explorar los resúmenes generados por Thot y participar en una trivia educativa diseñada para enriquecer su experiencia en el museo.

La solución propuesta consta de los siguientes componentes clave:

- **Extracción de Información:** Thot utiliza técnicas de NLP para analizar y extraer información relevante de los textos presentes en la sala Mexica, identificando conceptos clave, fechas importantes, personajes destacados y eventos significativos.
- **Síntesis de Información:** Con base en la información extraída, Thot genera resúmenes concisos y precisos que presentan los datos más relevantes de manera clara y accesible para los visitantes.
- **Interfaz de Usuario:** Thot cuenta con una interfaz intuitiva y amigable que permite a los usuarios explorar los resúmenes generados, acceder a información detallada y participar en la trivia interactiva.
- **Trivias Educativas:** Al finalizar la síntesis de la información, Thot presenta una trivia relacionada con los temas abordados en la sala Mexica, diseñada para que los visitantes pongan a prueba sus conocimientos y se diviertan aprendiendo.
- **Integración de Tecnologías Innovadoras:** Thot incorpora tecnologías avanzadas como NLP, IA y visualización de datos para ofrecer una experiencia educativa y cultural enriquecedora y atractiva para los visitantes del museo.

5. Alcance del Proyecto

El proyecto Thot tiene como objetivo desarrollar un sistema interactivo que utilice técnicas de procesamiento de lenguaje natural (NLP) y generación de imágenes mediante inteligencia artificial para mejorar la experiencia de los visitantes en la sala Mexica del Museo Nacional de Antropología. El sistema permitirá a los usuarios acceder a resúmenes concisos de la información exhibida y participar en actividades interactivas de trivia.

Funcionalidades Principales:

- **Sintetización automatizada de la información:** El sistema procesará textos históricos y descripciones de objetos presentes en la sala Mexica, generando resúmenes cortos que capturen los aspectos más relevantes para los visitantes.
- **Interacción mediante trivias:** Al final de cada sección de resumen, el sistema presentará una trivia interactiva, con preguntas relacionadas con los temas abordados en la síntesis de información, permitiendo la participación de grupos de visitantes (familias y amigos) en un entorno competitivo y educativo.
- **Sistema de ranking:** Los visitantes podrán acumular puntos en función de su desempeño en las trivias, generando un ranking que determinará al ganador de la actividad interactiva.

Límites del Proyecto:

- **Información limitada a la sala Mexica:** El sistema se centrará exclusivamente en los contenidos de la sala Mexica del Museo Nacional de Antropología y no abarcará otras salas o colecciones del museo.
- **Alcance temporal:** La implementación inicial del proyecto estará diseñada para operar durante un periodo de prueba, con la posibilidad de expandir su alcance en función de los resultados obtenidos y el feedback de los usuarios.

- Tecnologías específicas: El sistema utilizará técnicas de procesamiento de lenguaje natural para la síntesis de información y algoritmos de inteligencia artificial para la generación de imágenes. El desarrollo del proyecto se limitará a estas tecnologías y no incluirá hardware adicional más allá de lo necesario para la interacción básica con los usuarios (dispositivos móviles o terminales interactivas).

Excepciones:

- El sistema no incluirá interpretación de audios, videos o contenido multimedia, ni permitirá la interacción en tiempo real con guías u otros especialistas.

Resultados Esperados:

- Mejora en la experiencia educativa de los visitantes, permitiendo un acceso más ágil y comprensible a la información presentada en la sala Mexica.
- Incremento en la interacción entre grupos de visitantes a través de las trivias y la competencia amistosa, haciendo del aprendizaje una actividad compartida y lúdica.
- Recompensas personalizadas mediante IA que motiven a los usuarios a participar activamente y a mejorar su desempeño.

6. Análisis de Requerimientos

Requerimientos Funcionales

- Usuarios: El sistema será utilizado por un público diverso, lo que implica que la interfaz debe ser intuitiva y amigable para todas las edades. Además, el contenido debe ser adaptable para diferentes niveles de comprensión.
- Acceso a la información: Se destaca que el sistema será accesible desde un sitio web y funcionará a través de códigos QR, lo que permite una experiencia personalizada para cada usuario. Solo estará disponible en español inicialmente, pero es importante tener en cuenta una posible expansión multilingüe en el futuro.
- Interacción con el usuario: El proceso de interacción es simple: los usuarios escanean códigos QR para recolectar información de su interés y luego participar en una trivia. Este sistema es atractivo, ya que recompensa con imágenes personalizadas (fondo de pantalla) a aquellos usuarios que obtienen buenos puntajes.
- Generación de imágenes y trivia: Las imágenes generadas solo serán una recompensa al finalizar la trivia, lo que refuerza la motivación del usuario para interactuar con el sistema. El proceso está bien estructurado: el usuario escanea información, completa la trivia y, si su puntuación es alta, recibe la imagen personalizada.
- Personalización: Personalizar la información en función de la edad del usuario y ajustar la trivia a lo que ha escaneado asegura que la experiencia sea relevante para cada visitante. Además, mostrar las respuestas correctas con justificaciones al final de la trivia fomenta la curiosidad y el aprendizaje.

Requerimientos No Funcionales

- Actualización de contenido: Como la sala no cambia frecuentemente, no es necesario realizar actualizaciones frecuentes del contenido.
- Accesibilidad: Aunque inicialmente no hay consideraciones para accesibilidad, la futura inclusión de una opción de audio sería una mejora valiosa para personas con discapacidades visuales o problemas de lectura.
- Rendimiento: El sistema debe ser rápido, con tiempos de respuesta menores a 3 segundos para mantener la atención de los usuarios. También debe ser capaz de manejar hasta 1000 usuarios simultáneos, lo cual es importante en temporadas altas.
- Disponibilidad: El sistema solo estará disponible durante las horas de operación del museo, lo que puede reducir la demanda sobre la infraestructura, pero requiere una gestión eficiente durante los momentos pico.
- Seguridad y privacidad: No se almacenará información personal más allá de la edad del usuario y los datos se eliminarán al salir del museo.
- Escalabilidad: El sistema debe ser escalable para otras salas del museo, pero no se espera que la información cambie a menudo, lo que facilita la planificación de la infraestructura.
- Mantenimiento: Se planean mantenimientos cada 6 meses para la actualización de la base de datos y la corrección de errores, lo que parece suficiente dado el contenido estático del proyecto. Sin un equipo de soporte dedicado, la robustez del sistema será clave.
- Integración: La integración en la página web del museo es crucial para garantizar un fácil acceso. Además, no se requiere interactuar con otras tecnologías del museo, simplificando la implementación.

7. Calendarización de Actividades

Fase 1: Planificación y Análisis

Semana del 17 de agosto al 16 de septiembre de 2024 (trabajo previo a la primera entrega)

Análisis de Requerimientos Funcionales y No Funcionales (para entrega el 17 de septiembre):

- Reuniones con el equipo para recoger los requerimientos y validarlos (funcional y no funcional).
- Documentación de los requerimientos: identificar la funcionalidad clave de Thot y sus restricciones.
- Estimar el tiempo y costo del proyecto en base al alcance académico.

Cálculo del Proyecto:

- Crear el cronograma y definir métricas para las fases del proyecto.

Calendarización de Actividades:

- Crear el cronograma detallado usando una herramienta de gestión de proyectos (Trello, Asana, Gantt).

Fase 2: Análisis del Proyecto

Semana del 18 de septiembre al 23 de septiembre de 2024 (trabajo previo a la segunda entrega)

Análisis Inicial del Proyecto (para entrega el 24 de septiembre):

- **Modelo de Procesos:** Diagramar los procesos internos del sistema simulado (cómo se moverán los datos dentro del sistema).
- **Modelo de Datos:** Diseño preliminar del esquema de datos, incluyendo cómo se manejarán los datos de la trivia y los resúmenes.
- Validar y ajustar el análisis con el equipo.

Fase 3: Diseño del Sistema

Semana del 25 de septiembre al 30 de septiembre de 2024 (trabajo previo a la tercera entrega)

Análisis del Sistema (para entrega el 1 de octubre):

- **Modelo Funcional:** Documentar cómo los componentes del sistema interactuarán entre sí.
- **Modelo de Entrada/Salida:** Detallar el flujo de datos, describiendo la simulación de entrada de datos sobre la Sala Mexica y la salida a los usuarios.
- **Diseño de Interfaces:** Prototipos iniciales para el sistema interactivo.

Semana del 2 de octubre al 7 de octubre de 2024 (trabajo previo a la cuarta entrega)

Diseño Inicial del Sistema (para entrega el 8 de octubre):

- **Diseño Operativo:** Planificar la arquitectura del sistema simulado, considerando la interacción entre módulos (NLP, trivia, interfaz).
- **Diseño de Interfaces:** Crear wireframes para las pantallas del sistema interactivo.

Semana del 9 de octubre al 14 de octubre de 2024 (trabajo previo a la quinta entrega)

Diseño Completo del Sistema (para entrega el 15 de octubre):

- **Diseño Funcional:** Detallar cómo funcionarán los módulos clave como el procesamiento de lenguaje natural y el sistema de trivia.
- **Diseño Operativo:** Optimizar la arquitectura del sistema para su simulación en el entorno académico.

Fase 4: Implementación y Desarrollo

Semana del 16 de octubre al 30 de octubre de 2024

Desarrollo del Módulo de Procesamiento de Lenguaje Natural:

- Implementar el sistema para la extracción y resumen de información usando datos simulados.
- Pruebas iniciales del procesamiento de textos.

Semana del 1 al 14 de noviembre de 2024

Desarrollo del Módulo de Trivia:

- Implementar el generador de preguntas de trivia y el sistema de ranking.
- Pruebas unitarias para verificar la funcionalidad de la trivia.

Semana del 15 al 30 de noviembre de 2024

Desarrollo del Sistema de Interfaz:

- Implementar la interfaz de usuario y conectar los módulos de resumen y trivia.
- Pruebas iniciales del sistema en un entorno simulado.

Fase 5: Integración, Pruebas y Validación

Semana del 1 al 14 de diciembre de 2024

Pruebas de Integración:

- Integrar todos los módulos y realizar pruebas para verificar su correcta interacción.

Pruebas Funcionales:

- Verificar el funcionamiento completo del sistema y realizar ajustes según los resultados de las pruebas.

Semana del 15 al 23 de diciembre de 2024

Pruebas en Entorno Simulado:

- Ejecutar pruebas de usuario en un entorno completamente simulado, asegurando que todo funcione según lo planeado.
- Ajustes finales en función de los resultados obtenidos en las pruebas.

Fase 6: Entrega Final y Presentación

Semana del 24 al 30 de diciembre de 2024

Optimización Final:

- Mejorar el rendimiento y corregir errores restantes.
- Finalizar la documentación del proyecto académico (manual técnico y de usuario).

Semana del 31 de diciembre de 2024 al 7 de enero de 2025

Preparación para la Entrega:

- Preparar la presentación del proyecto y demostrar el sistema funcionando en un entorno simulado.

10 de enero de 2025

Entrega Final del Proyecto:

- Presentación formal del sistema Thot como parte del proyecto escolar.

8. Cálculo del Proyecto

8.1 Métricas de Desempeño del Proyecto Thot

Tiempo estimado vs. tiempo real: Se espera que el desarrollo del proyecto Thot concluya en un plazo de seis meses, considerando que el equipo de desarrollo estará compuesto por tres personas trabajando a tiempo parcial. El tiempo estimado se ha distribuido en varias fases, que incluyen la recolección de datos, el desarrollo de los algoritmos de Procesamiento de Lenguaje Natural (NLP), la implementación del sistema de trivia y las pruebas finales. Esta métrica compara el tiempo planeado para cada fase con el tiempo real que se empleó para completarlas. Si se encuentran desviaciones significativas, será necesario hacer ajustes en el calendario para asegurar que el proyecto se mantenga dentro del plazo.

Porcentaje de avance: El proyecto se ha dividido en cinco fases principales: diseño del sistema, desarrollo de la síntesis de información, desarrollo del sistema de trivia, pruebas y optimización final. El porcentaje de avance indicará qué proporción del proyecto ha sido completada en cada una de estas fases, en comparación con el total planificado. Por ejemplo, al finalizar la primera fase, se espera que aproximadamente el 20% del proyecto esté completo. Esta métrica permitirá monitorear el progreso general del desarrollo.

Número de tareas completadas vs. pendientes: El proyecto está organizado en tareas específicas, como la implementación de los resúmenes basados en NLP y la integración del sistema con el sitio web del museo. Esta métrica mide cuántas tareas han sido completadas en comparación con las que están pendientes. Se espera que el equipo realice un seguimiento semanal del número de tareas finalizadas, lo que permitirá mantener el control del proyecto y tomar decisiones a tiempo si se detecta algún retraso.

8.2 Métricas de Calidad del Proyecto Thot

Errores detectados: Durante el desarrollo del proyecto Thot, se realizarán al menos tres ciclos de pruebas: pruebas internas por parte del equipo de desarrollo, pruebas de usuario con un grupo reducido de visitantes seleccionados, y pruebas en condiciones reales en el museo. Esta métrica cuantifica el número de errores detectados en cada ciclo de pruebas, como fallos en la generación de resúmenes o errores en el sistema de trivia. El objetivo es reducir el número de errores críticos a menos de cinco en la fase final de pruebas, lo cual indicará que el sistema es estable y está listo para ser implementado.

Satisfacción del usuario: Una vez que el sistema esté en funcionamiento en el museo, se recogerán opiniones de los usuarios a través de encuestas y formularios disponibles en el sitio web del museo. La meta es alcanzar un nivel de satisfacción del usuario del 80 % o más, evaluando aspectos como la facilidad de uso, la claridad de los resúmenes proporcionados y el entretenimiento del sistema de trivia. Esta métrica ayudará a asegurar que el sistema Thot cumpla con las expectativas educativas y de interacción de los visitantes.

Tasas de revisión y corrección: Durante el proceso de desarrollo y pruebas, el equipo revisará las funcionalidades clave del sistema para detectar y corregir posibles fallos. Esta métrica mide cuántas veces se realizaron revisiones y correcciones en el sistema. Se espera que después del primer ciclo de pruebas, al menos el 60 % de las correcciones sean menores, lo que indicaría que el sistema está bien estructurado y que las correcciones significativas se realizaron en las fases iniciales.

8.3 Métricas de Costos del Proyecto Thot

Costo planeado vs. costo real: El presupuesto total del proyecto Thot se estima en \$100,000 MXN, incluyendo costos de desarrollo, pruebas y mantenimiento. Esta métrica compara los costos reales incurridos en cada fase del proyecto con los costos previstos. Por ejemplo, si se estimaron \$30,000 MXN para la fase de pruebas, pero se gastaron \$35,000 MXN, será necesario ajustar el presupuesto de las fases siguientes para evitar que el costo total exceda el presupuesto estimado.

Costo por tarea/unidad: Cada tarea o funcionalidad del sistema tendrá un costo asociado, ya sea en términos de horas de desarrollo o en costos directos (como licencias de software o infraestructura). Por ejemplo, se estima que el desarrollo del sistema de trivia costará \$10,000 MXN en total, considerando tanto la implementación como las pruebas. Esta métrica ayuda a identificar qué tareas están consumiendo más recursos de lo previsto y permite realizar ajustes para mantener el control financiero del proyecto.

Desviación presupuestaria: Esta métrica mide la diferencia entre el costo presupuestado y el costo real. Se ha establecido que la desviación máxima permitida será del 10 % sobre el presupuesto total del proyecto. Si la desviación supera este límite, se deberá justificar y tomar medidas para corregir el curso del proyecto, ya que la meta es evitar cualquier sobrecoste significativo que impacte el desarrollo o la implementación.

8.4 Métricas de Productividad del Proyecto Thot

Tasa de productividad por equipo o individuo: Se espera que cada miembro del equipo de desarrollo complete un número determinado de tareas cada semana. Esta métrica mide cuántas tareas fueron finalizadas en comparación con el tiempo invertido. Por ejemplo, si a un desarrollador se le asignaron 10 tareas y completó solo 5 en una semana, se evaluará si la carga de trabajo es adecuada o si existen barreras que están afectando su productividad. Esta métrica permite medir la eficiencia individual y del equipo en general.

Velocidad de desarrollo: La velocidad de desarrollo se medirá en cada fase del proyecto, indicando cuántas funcionalidades fueron implementadas dentro del plazo establecido. Por ejemplo, la fase de implementación del sistema de resúmenes y trivias tiene un tiempo estimado de dos meses, y esta métrica monitoreará si el equipo ha logrado completar el 80 % de las funcionalidades en ese tiempo. Una velocidad de desarrollo consistente asegura que el proyecto se mantenga dentro del plazo.

8.5 Métricas de Riesgo del Proyecto Thot

Número de riesgos identificados: Al comienzo del proyecto se identificaron varios riesgos potenciales, como la dificultad para integrar el sistema con la infraestructura tecnológica del museo o posibles problemas con la precisión del procesamiento de lenguaje natural. Esta métrica cuantifica cuántos riesgos se han identificado durante el desarrollo del sistema. Se espera identificar los riesgos clave en las fases iniciales del proyecto para que puedan ser mitigados antes de causar problemas mayores.

Impacto de riesgos mitigados: Esta métrica evalúa la efectividad de las medidas implementadas para mitigar los riesgos. Por ejemplo, para evitar sobrecargas en los servidores del museo, se podría integrar un sistema de caché. El impacto de estas medidas se medirá por la reducción del riesgo, con el objetivo de disminuir la probabilidad de fallos en un 80 %. Esta métrica es esencial para asegurar que el proyecto avance sin interrupciones graves y que los riesgos sean gestionados adecuadamente.