

Plan de desarrollo de protocolo para PEI 2015

Plan de desarrollo de protocolo para PEI 2015 - ITAM

Raúl Benítez peña

Contenidos

Plan de desarrollo de protocolo para PEI 2015.....	3
Requerimientos hacia las IES.....	3
Nombre del proyecto.....	3
Actividades necesarias y sugeridas.....	3
Descripción de las actividades.....	4
Costo aproximado.....	5
Cronograma.....	6
Participantes.....	7
Entregables.....	8
Resultados esperados.....	9
Contactos.....	10
Técnico.....	10
Administrativo.....	10

Plan de desarrollo de protocolo para PEI 2015

Derivado de las reuniones sostenidas los días 6 y 9 de Octubre, les solicitamos por favor llenen el siguiente documento, el cual recaba la información necesaria que servirá como base de las propuestas presentadas al CONACYT.

Este documento deberá ser replicado según el número de propuestas de proyectos que realicen.

Requerimientos hacia las IES

Mediante la recopilación la información requerida a continuación podremos tener un esquema más claro sobre el desarrollo y requerimientos del PEI por lo que les pedimos nos apoyen en su llenado.

Nombre del proyecto

"Desarrollo de un sistema de publicación de biblioteca de arte mexicano."

En particular nos enfocaremos en resolver los siguientes puntos:

- Las herramientas de similitud: Las cuales permiten agrupar de forma automatizada los contenidos acorde a sus temáticas.
- Las herramientas back-end para realizar búsquedas: Acorde a una búsqueda introducida estos módulos son capaces de detectar temas similares a los solicitados por el usuario.
- Las herramientas de resumen: Que puedan trabajar acorde a resultados de búsqueda y por unidades de contenidos (libros, capítulos de libros o artículos).
- Las herramientas de clasificación: Capaces de clasificar documentos de manera automática en categorías existentes.

Actividades necesarias y sugeridas

NOTA: La lista que a continuación se presenta, es una vista de alto nivel de las actividades a realizarse. Se determinaron a partir de la información con la que actualmente se cuenta.

1. Adquisición del estado de los datos y metadatos con los que cuenta Eficiencia Informativa en la biblioteca de Arte Mexicano.
2. Recolección bibliográfica, artículos académicos y de la industria para establecer el estado del arte.
3. Adquisición de conocimiento de arquitectura de datos, tecnología y de sistemas que actualmente posee Eficiencia Informativa.
4. Estructura de los datos que actualmente posee Eficiencia Informativa, así como de los procesos que tienen para generarlos de manera continua (formatos, flujos, tablas, motores, drivers, periodicidad, tamaño actual, crecimiento, etc.)
5. Solicitud de datos y recepción de los mismos.
6. Establecimiento de arquitectura del equipo de Ciencia de Datos (DS). Construcción y montado en Amazon AWS. Conexiones y accesos. Establecer niveles de seguridad.
7. ETL (*scripts* de extracción, transformación y carga) de datos proporcionados por Eficiencia Informativa a la arquitectura de DS.
8. Listado de algoritmos a implementar.
9. Implementación de algoritmos.
10. Visualización y métricas de resultados.
11. Establecimiento de proceso continuo.
12. Sugerencias de implementación.

Descripción de las actividades

ID	Actividad y descripción	Justificación	Tiempo (días)	Fecha inicio	Entregable
1	Conocimiento de datos.	Adquirir el conocimiento acerca de los datos y metadatos con los que cuenta Eficiencia Informativa en la biblioteca de Arte Mexicano	30 días	01/01/2015	
2	Recolección bibliográfica, artículos académicos y de la industria para establecer el estado del arte.	Adquirir conocimiento de proyectos similares e identificar las mejores prácticas.	28 días	01/02/2015	
3	Adquisición de conocimiento de arquitectura de datos, tecnología y de sistemas.	Definición de las tecnologías a utilizar (paquetería de software) e infraestructura necesaria para recopilar la información, analizarla, integrar el trabajo anterior y escalarla (desarrollo, prueba y producción).	15 días	01/03/2015	
4	Estructura de datos: definir la organización y la estructura del sistema de almacenamiento de los datos	Estructurar la información para el proyecto a desarrollar con el fin de mejorar la visualización.	30 días	15/03/2015	
5	Recopilación de datos.	Insumo principal sobre el que se va a trabajar , por lo que este paso es de suma importancia para las actividades subsecuentes.	30 días	15/04/2015	
6	Establecimiento de arquitectura de DS. Determinar e implementar la infraestructura para el desarrollo del proyecto acorde a las necesidades que vayan surgiendo.	Asegurar el funcionamiento de la arquitectura necesaria para la implementación del proyecto y para que sea consistente con una futura implementación / continuación	30 días	15/05/2015	
7	Extracción , transformación y carga de datos (ETL) a la arquitectura de DS.	Procesamiento de los datos disponibles e incorporación de nuevas bases de datos que enriquezcan el análisis	45 días	15/06/2015	
8	Listado de algoritmos a implementar. Escritura del código necesario para el modelado y análisis del proyecto	Selección de los algoritmos a generar para el estudio	45 días	30/07/2015	
9	Evaluación de algoritmos	Comparación y clasificación de la metodología utilizada para elegir el algoritmo con mejor desempeño en operación	30 días	30/08/2015	
10	Visualización y métricas de resultados	Generación de un entregable útil que produzca valor agregado por medio de herramientas visuales	30 días	30/09/2015	
11	Establecimiento de procesos continuos	Asegurar que la implementación permita dar seguimiento al proyecto	30 días	30/10/2015	
12	Sugerencias de implementación	Presentar propuestas viables que generen sinergias entre los diferentes componentes del proyecto	30 días	30/11/2015	

Costo aproximado

Indique el precio aproximado del proyecto.

Concepto	Unidades	Precio Unitario	Total
Cómputo y almacenamiento (fijo)	- 2 máquinas r3.2xlarge (o similar) - 1 máquina m3.xlarge (o similar) - s3 o Mongoddb o Dynamo DB	- 1,179 USD (anual) - 590 USD (anual) - entre 400 y 2400 USD	69,524 MXP
Cómputo y almacenamiento (on demand)	EMR	entre 94 y 2,367 USD	30,771 MXP
Becas de manutención	4	6,000 MXN	288,000 MXP
Créditos académicos	180	3,015 MXN	542,700 MXP
Equipo de Cómputo	4	2,000 USD	108,234 MXP
Conferencias	2	80,000 MXP	160,000 MXP
Visitantes	2	80,000 MXP	160,000 MXP
TOTAL			1,359,229 MXP

Cronograma

Participantes

ID	Nombre del participante	Grado de estudios	CVU	ID actividades asignadas
1	Carlos Eduardo Petricioli Araiza	Lic. en Matemáticas Aplicadas / Lic. en Economía		
2	Andrés Villaseñor Lechuga	Licenciatura en matemáticas		
3	Carlos Ayala Raggi	Licenciatura en Computación		
4	Jose Antonio Mancilla Chatu	Lic. en Ingeniería Industrial		
5	Alejandra Barrera Ramírez	Dra. En Ingeniería	242070	
6	Adolfo De Unánue Tiscareño	Dr. en Física Teórica	36821	

Nota: Los participantes numerados del 1-4 se encuentran estudiando la maestría de Ciencia de Datos en el ITAM.

Entregables

ID	Entregable	Fecha de entrega
1	Definición de tecnologías, infraestructura básica y arquitectura. Definición del pipeline completo del proyecto.	15/06/2015
2	Integración de los datos disponibles, documentación de la metodología y procesos empleados. Generación de estándares, metodologías y ETLs para la inserción de nuevos documentos.	30/07/2015
3	Documentación del listado de algoritmos, criterio de selección de éstos, entrenamiento y validación de los prototipos. Implementación (en conjuntos de entrenamiento y prueba), automatización y medidas de desempeño. Prototipos de presentación de resultados.	30/09/2015
4	Presentación del desempeño de resultados obtenidos con los modelos finales. Visualización de estos en formato digital.	30/11/2015
5	Despliegue, automatización de procesos faltantes y documentación final.	31/12/2015

Resultados esperados

El proyecto “Biblioteca de Arte Mexicano” busca permitir a usuarios explorar de una manera fácil, intuitiva la historia del arte mexicano. Para lograrlo se proponen la generación de algoritmos para resumir contenido, extraer entidades y sus relaciones, análisis de tópicos en periodos determinados y relación de las obras por contenido.

Esperamos con este proyecto contribuya a la generación de conocimiento en áreas de alta tecnología de recursos humanos pertenecientes al programa de Maestría en Ciencias de Datos del Instituto Tecnológico Autónomo de México (ITAM).

Contactos

Técnico

Dr. Adolfo J. De Unánue Tiscareño

adolfo.deunanue@itam.mx

Teléfono 56.28.4000 ext 4052

Administrativo

Indicar nombre completo, teléfono y extensión y número celular de la persona que quedará como encargada de la parte administrativa del proyecto.