

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE OCCIDENTE

Departamento de Electrónica, Sistemas e Informática

Desarrollo Tecnológico y Generación de Riqueza Sustentable

PROYECTO DE APLICACIÓN PROFESIONAL (PAP)



ITESO
Universidad Jesuita
de Guadalajara

**4B03 PAP PROGRAMA DE DESARROLLO DE SOFTWARE DE ALTO
DESEMPEÑO I**

ITESO, Tlaquepaque.

PRESENTA

Ingeniería en Sistemas Computacionales, Carmen Rivera Jiménez,

Profesor PAP: Act. Juan Manuel Islas Espinoza, PMP®

Tlaquepaque, Jalisco, mayo de 2018.

ÍNDICE

Contenido

REPORTE PAP	3
Presentación Institucional de los Proyectos de Aplicación Profesional	3
Resumen	4
1. Introducción	5
1.1. Objetivos	5
1.2. Justificación.....	5
1.3 Antecedentes	5
1.4. Contexto.....	6
2. Desarrollo	7
2.1. Sustento teórico y metodológico	7
2.2. Planeación y seguimiento del proyecto	7
3. Resultados del trabajo profesional.....	8
4. Reflexiones del alumno o alumnos sobre sus aprendizajes, las implicaciones éticas y los aportes sociales del proyecto	11
5. Conclusiones	13
6. Bibliografía.....	15
Anexos (en caso de ser necesarios).....	15

REPORTE PAP

Presentación Institucional de los Proyectos de Aplicación Profesional

Los Proyectos de Aplicación Profesional (PAP) son una modalidad educativa del ITESO en la que el estudiante aplica sus saberes y competencias socio-profesionales para el desarrollo de un proyecto que plantea soluciones a problemas de entornos reales. Su espíritu está dirigido para que el estudiante ejerza su profesión mediante una perspectiva ética y socialmente responsable.

A través de las actividades realizadas en el PAP, se acreditan el servicio social y la opción terminal. Así, en este reporte se documentan las actividades que tuvieron lugar durante el desarrollo del proyecto, sus incidencias en el entorno, y las reflexiones y aprendizajes profesionales que el estudiante desarrolló en el transcurso de su labor.

Resumen

El proyecto descrito a continuación tiene la finalidad de generar bases de conocimiento y una guía escrita sobre técnicas de minería de datos apoyándose de un lenguaje de programación de alto nivel y una herramienta de procesamiento de lenguaje natural.

El procesamiento de lenguaje natural es un campo de las ciencias de la computación, inteligencia artificial y lingüística que estudia las interacciones entre las computadoras y el lenguaje humano

El proyecto actualmente está dividido en dos etapas, una de investigación y otra de aplicación. Durante esta etapa de investigación, se tomó como caso de estudio el tema de las elecciones para la presidencia de la república mexicana de 2018. Se categorizó el contenido de textos no estructurados como son las publicaciones de redes sociales (Facebook y Twitter), para ello se empleó una serie de algoritmos y funciones para de esta manera analizar tendencias sociológicas actuales.

La metodología utilizada para este proyecto de investigación es la de tipo experimental en la cual se probaron distintos algoritmos hasta tener el resultado esperado del análisis de textos.

A continuación, se describirá más a detalle el desarrollo del proyecto en cuestión.

1. Introducción

1.1. Objetivos

Generar conocimiento sobre herramientas, técnicas y métodos para obtener, modelar y explotar conocimiento registrado en textos no estructurados. Esta investigación tendría un impacto local en la universidad para apoyar a distintas carreras a obtener información no digitalizada como tal o ubicada en redes sociales y diversas páginas de internet.

Para esto pretendo obtener aprendizaje sobre un lenguaje de programación de alto nivel como lo es Python además de una herramienta de procesamiento de lenguaje natural libre como NLTK.

1.2. Justificación

Dado que es un proyecto de investigación, debo ser responsable, organizada y autodidacta para alcanzar a obtener los niveles de conocimiento adecuados para llevar a cabo el proyecto. A pesar de que balancear el tiempo entre otras actividades y el proyecto puede llegar a ser complicado, considero que los aprendizajes que puedo llegar a obtener serán de mucho provecho en mi vida profesional.

Creo que estar estudiando sobre diversos temas al mismo tiempo que se desarrolla el proyecto, hace que pueda entender mejor el contexto en el cual se desarrolla el mismo. Por otro lado, estos aprendizajes me llevarán a completar el proyecto en tiempo y forma.

1.3 Antecedentes

Una de las cosas que me agradó mucho del proyecto fue tener la oportunidad de aprender un lenguaje de programación nuevo para mí y conocer herramientas de procesamiento de lenguaje natural, así como aprender sobre el concepto de minería de datos.

Dado que el proyecto es de investigación me llamó mucho la atención el poder distribuir mis tiempos para trabajar en el con mayor facilidad y tener la confianza por parte del encargado del proyecto de que realizaré las tareas que me encomienda

Considero que este proyecto me ayuda a ir dando cierre a mi formación escolar puesto que aplico mis conocimientos de programación y distintas tecnologías, en un proyecto real con necesidades reales que si se realiza adecuadamente podrá ayudar a más personas dentro de la institución.

El proyecto se encuentra actualmente en el departamento de Electrónica, Sistemas e Informática del Instituto Tecnológico de Estudios Superiores de Occidente. Ésta es una institución educativa privada con 60 años de antigüedad.

1.4. Contexto

En este proyecto se planea obtener y analizar información ubicada en textos no estructurados del internet para investigar tendencias sociales con respecto a las elecciones del 2018. Esto se logra por medio de API's de acceso a perfiles públicos de Twitter o Facebook. Posteriormente se analizarán los textos mediante herramientas de procesamiento de lenguaje natural.

Los entregables a realizar en esta etapa del proyecto son: un reporte técnico con bibliografía y análisis de los resultados obtenidos del caso de estudio en cuestión.

Este proyecto se divide en 2 etapas. La primera etapa que es en la que se basa este proyecto, es la investigación. Se generará conocimiento y scripts para extracción y análisis de datos con un caso de estudio como son las elecciones del 2018.

La segunda etapa pretende desarrollar una aplicación, a las que puedan tener acceso los estudiantes de otras carreras para estudiar tendencias sociales de distinta índole sin necesidad de tener conocimientos tecnológicos avanzados.

En este proyecto estaré desarrollando el rol de asistente de investigación, en el cual estaré encargada de investigar técnicas de procesamiento de lenguaje humano y nuevas tecnologías sobre análisis lingüístico en línea.

La minería de datos es una herramienta muy útil actualmente en empresas para descubrir patrones en grandes volúmenes de datos, por lo que el conocimiento adquirido durante este proyecto tendrá gran utilidad para mí a futuro.

2. Desarrollo

2.1. Sustento teórico y metodológico

La metodología que se utiliza es el método científico puesto que es un proyecto de investigación. Los pasos que utilizo son:

1. Realizar hipótesis
2. Experimentación
3. Teoría

Para este proyecto se utilizará esta metodología aplicada al uso de una herramienta de procesamiento de lenguaje natural llamada NLTK (Natural Language Tool Kit) con la cual se realizarán experimentos para probar ciertas hipótesis.

Dichas hipótesis serán probadas o refutadas por medio de análisis de sentimientos. Las hipótesis planteadas serán referentes a las opiniones de los usuarios de Twitter y Facebook referente a las elecciones presidenciales de México 2018.

2.2. Planeación y seguimiento del proyecto

- Descripción del proyecto

Para realizar este proyecto se tiene acordada una sesión semanal de revisión con el asesor del proyecto para dar seguimiento. En estas sesiones el asesor me asigna tareas que debo cumplir para la próxima sesión. El reporte técnico se comenzará a redactar cuando se empiecen a tener los primeros resultados de los experimentos.

Al final del proyecto se debe tener un nivel medio de conocimiento sobre el lenguaje Python y nivel medio también sobre la herramienta NLTK. Para ello deberé estar tomando un curso en línea sobre programación en Python para ir alcanzando el nivel requerido por el proyecto.

No.	Competencia	Nivel Requerido al inicio	Nivel Adquirido al Inicio	Nivel Objetivo al final PAP	Prior
1	Conocimientos sobre NLTK	2	1	2	B
2	Programación en Python	3	1	3	B
3	Uso de controladores de versiones (GitHub)	1	2	2	A
4	Conocimiento básico de JSON	2	1	2	B
5	Comunicación Oral y Escrita en inglés	3	3	3	A

- Plan de trabajo

Plan de Actividades

No.	Actividad Educativa	Tipo Actividad	Prereq	Total Hrs	Fecha Inicio	Fecha Termina	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Obj	
1	Organización	Autoestudio	N/A	5	18-ene	25-ene																		3
2	Autodidacta	Autoestudio	N/A	40	18-ene	22-feb																		3
3	Conocimiento de inglés	Autoestudio	N/A	0	N/A	N/A																		3
3.1	Lectura en inglés	Autoestudio	N/A	0	N/A	N/A																		3
3.2	Escritura en inglés	Autoestudio	N/A	0	N/A	N/A																		3
4	Ambientes de desarrollo	Autoestudio	N/A	0	N/A	N/A																		3
5	Programación en Python	Curso en línea	N/A	30	22-ene	02-mar																		2
5.1	Variables, tipos de datos y operadores	Curso en línea	N/A	10	22-ene	02-feb																		2
5.2	Condicionales, bucles y funciones	Curso en línea	5.1	10	05-feb	16-feb																		2
5.3	Diccionarios, módulos y paquetes	Curso en línea	5.2	10	19-feb	02-mar																		2
6	Controlador de versiones	Autoestudio	N/A	0	N/A	N/A																		2
7	NLTK (Natural Language Tool Kit)	Autoestudio	N/A	40	05-feb	06-abr																		2
7.1	Uso de NLTK mediante python	Autoestudio	5.1	10	05-feb	16-feb																		2
7.2	Procesamiento de texto	Autoestudio	7.1	10	05-feb	02-mar																		2
7.3	Categorización de textos	Autoestudio	7.2	10	05-mar	16-mar																		2
7.4	Extracción de información de textos	Autoestudio	7.3	10	19-mar	30-mar																		2
8	Elaboración de reporte técnico	Autoestudio	7.4	40	01-abr	11-may																		2

- Comunicaciones

Emisor	Mensaje	Receptor	Medio	Frecuencia
Carmen	Documento	Profesor PAP	Moodle	Quincenal
Carmen	Duda o aclaración	Asesor de proyecto	Email	Semanal
Carmen	Reporte	Asesor del proyecto	Email	Semanal

- Calidad

Quién Entrega	Qué Entrega (Entregable)	A Quién recibe o Inspecciona	(Criterios de Aceptación)	Siguiente paso. Cómo Autoriza?
Carmen Rivera	Resultado de experimento	Víctor Hugo Zaldívar	Que sea coherente	Menciona siguiente experimento
Carmen Rivera	Código funcional	Víctor Hugo Zaldívar	Que funcione de acuerdo con lo descrito	Pide investigue otro tema

- Equipo de Trabajo

Rol	Responsabilidad	Nombre
Asistente de investigación	Investigar y realizar los experimentos solicitados por el asesor.	Carmen Rivera
Líder del proyecto	Asignar tareas y resolver dudas.	Víctor Hugo Zaldívar

- Seguimiento

Cada semana tengo una reunión con el asesor del proyecto para resolución de dudas. Además, se acuerdan las actividades y tareas que se realizarán para la siguiente semana. Adicional a esto

si tengo alguna duda antes de la reunión de revisión, mando un correo a mi asesor y dependiendo de la urgencia de la duda se puede agendar otra sesión de asesoría.

Para la elaboración del reporte PAP tengo una clase semanal con el profesor PAP quien me explica cómo debo ir llenando los capítulos del reporte. Se sube el capítulo terminado a Moodle y si hay algún error se tiene una sesión de asesoría para corregir el reporte.

Al finalizar el PAP (*sesión 16 del semestre*), entregaré el Reporte Final PAP y expondré presencialmente un resumen del Reporte Final PAP ante el grupo y autoridades de la Coordinación.

Adicional a esto, tendré una sesión de cierre con el asesor de mi proyecto quien me dará retroalimentación sobre mi desempeño durante todo el proyecto.

3. Resultados del trabajo profesional

3.1 Productos obtenidos

Dado que el proyecto realizado era de investigación, uno de los entregables finales fue un reporte técnico donde describí paso a paso las acciones que tuve que realizar para mis experimentos, así como los hallazgos obtenidos del caso de estudio utilizado.

Además, entregué a mi asesor el código fuente de todos los Scripts en lenguaje Python que realicé, para tener un banco de Scripts para futuros proyectos referentes a minería de datos.

3.2 Estimación del impacto

A partir de este proyecto se podrán realizar más investigaciones referentes al uso de minería de datos en redes sociales. Su uso es muy importante hoy en día debido a que tiene muchísimas áreas de aplicación tales como mercadotecnia, toma de decisiones empresariales, estudios sociológicos, análisis de tendencias entre otras.

Por otro lado, a partir de la información ya recabada es muy factible el construir una aplicación ya sea móvil o web que se apoye de estos scripts ya desarrollados, para así proporcionar a estudiantes de otras carreras una herramienta que les permita extraer información útil de redes sociales para distintos proyectos con mayor facilidad y eficiencia. Esta herramienta podría ser utilizada a nivel ITESO por lo menos en las carreras de mercadotecnia, administración y psicología.

4. Reflexiones del alumno o alumnos sobre sus aprendizajes, las implicaciones éticas y los aportes sociales del proyecto.

4.1 Aprendizajes profesionales

Gracias a este proyecto mejoré mis habilidades de auto estudio al tener que investigar muchos conceptos completamente nuevos por mi cuenta en diversas fuentes, me parece que esto será de mucho provecho en mi vida laboral porque la tecnología y lenguajes de programación están cambiando constantemente así que necesitaré estar aprendiendo constantemente.

Me agradó mucho el aprender sobre procesamiento de lenguaje natural ya que tiene aplicaciones en muchas áreas como son mercadotecnia, finanzas, y sociología entre otras para analizar tendencias para beneficiar a las organizaciones.

Otra de las cosas que aprendí fue a programar en Python y me gustó mucho el manipular un lenguaje de alto nivel como tal. Considero que puede tener muchos usos en mi vida profesional debido a todas las aplicaciones que tienen.

4.2 Aprendizajes sociales

El conocimiento que generé al desarrollar este proyecto podrá ayudar a muchos estudiantes de la universidad a recopilar información de distintas fuentes con mucha mayor eficiencia. Para esto, se planea que para siguientes proyectos relacionados con procesamiento de lenguaje natural puedan desarrollar una interfaz con la que sea más fácil para los usuarios extraer información de diversas fuentes para su análisis.

Si este proyecto se desenvuelve de manera exitosa en el campus, se podría incluso lanzar una aplicación a mayor escala para poner sus herramientas al alcance de muchas personas y no sólo de la universidad.

Después de realizar este proyecto considero que será mucho más fácil para mi participar en otros proyectos de investigación ya que logré aprender de qué manera debo dividir mis tiempos para lograr los resultados esperados y en qué tipo de fuentes puedo confiar para obtener información.

4.3 Aprendizajes éticos

A la hora de realizar este proyecto tuve que poner mucho esfuerzo en tratar de dar a cada materia y actividad el tiempo necesario para realizarla de manera satisfactoria. Pero a final de cuentas opté por dedicar el tiempo acordado en cada cosa y al final los resultados fueron satisfactorios en ambas partes.

Otro aprendizaje que me dejó el proyecto fue mostrar responsabilidad y respeto. A la hora de tener sesiones de dudas o revisiones con el asesor de mi proyecto, fue muy importante que

hubiera respeto de ambas partes para no disponer del tiempo de la otra persona si llegaba a suceder algún inconveniente o impedimento para tener la sesión. Considero que esto es muy valioso en la vida laboral ya que como profesional habrá juntas y eventos a los que tendré que asistir y deberé asistir en tiempo y forma a ellos.

4.4 Aprendizajes en lo personal

Para mi este proyecto me ayudó mucho a desenvolverme de manera independiente en un proyecto. A pesar de tener un guía y asesor, me di cuenta de que siempre debo investigar lo más posible para evitar estar preguntando en todo momento ya que no sólo aprenderé sobre el tema que busco, sino que obtendré más experiencia en como buscar información ya sea en internet, libros o diversas fuentes de información que pudiera tener una organización.

En lo personal me agradó muchísimo el tema de procesamiento de lenguaje natural y creo que seguiré investigando al respecto sobre el tema, e incluso a considerar ofertas de trabajo que involucren minería de datos.

5. Conclusiones

En ocasiones mi asesor no podía tener las sesiones de revisión por lo que se alargaban las esperas para preguntar dudas o simplemente saber qué más hacer. Después de que ocurrió la primera vez, me di cuenta de que debía ser previsor y preguntar de antemano qué otra cosa podría ir aventajando si por algo no tendríamos revisiones.

Al principio del proyecto solo se tenía contemplado realizar minería con Twitter, pero me pareció que tanto en Twitter como en Facebook podríamos obtener mucha información útil así que se lo mencione a mi asesor, este respondió que sería un poco más complicado, pero accedió por lo que vi que a pesar de que yo era asistente de investigación yo podía proponer y exponer mis ideas y el las tomaría en cuenta.

Me gustó mucho el proyecto ya que el tema es muy interesante y tiene campos de aplicación muy diversos. Además, el investigar por mi cuenta me dejó muchas competencias como auto estudio, responsabilidad y organización.

Considero que la relación entre trabajo, esfuerzo y resultados fue muy buena porque a pesar de que había semanas que sentía que tenía muchas actividades y no lograría terminarlas, se compensaba con otras en que eran más sencillas y podía dedicarme a otras actividades.

5.1 Tareas Aprendidas

Gracias a este proyecto tuve aprendizajes en varios ámbitos y hay cosas que debo repetir y mejorar y otras que deben cambiarse.

Entre las actitudes que favorecieron a que el proyecto saliera bien están el ser responsable y puntual en las sesiones con el asesor, ser organizada para dividir bien mis tiempos entre todas mis otras actividades y tener calidad en los entregables. Por parte del asesor, me gustó que me dejara libertad sobre investigar temas que me interesaran puesto que esto favoreció que me sintiera más involucrada con el proyecto.

Las cosas que pudiera mejorar para futuros proyectos son el no tener pena de preguntar mis dudas al momento que me surgen. En las primeras semanas tuve ciertas preguntas y esperaba a la siguiente sesión para pedir ayuda. Por parte del asesor sería tener un mejor manejo del calendario porque en ocasiones no podíamos tener la sesión de asesoría debido a que tenía otras juntas y compromisos.

Considero que el documentar el proyecto es muy importante porque así puedes tener más en mente el avance que vas teniendo y el alcance esperado para el final del mismo. Me gustó trabajar en el ambiente de investigación puesto a que es muy flexible y me permitió explorar diferentes tecnologías que eran nuevas para mí.

Pienso que una vez que termine mis estudios, deberé buscar un trabajo en el que esté aprendiendo constantemente sobre nuevas tecnologías y donde rete mis habilidades para investigar sobre ellas.

6. Bibliografía

1 Bird, S., Klein, E., Loper, E. Analyzing Text with the Natural Language Toolkit. NLTK Org.
Sitio web: <http://www.nltk.org/book/>

Anexos (en caso de ser necesarios)

[Es todo lo que soporte el documento; que no es necesario añadirlo en el cuerpo del reporte, pero que sirve para su mejor comprensión: materiales elaborados, bitácoras, fichas, fotografías, mapas, etc.]