



DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA

CARRERA DE INGENIERÍA ELECTRÓNICA, EN
TELECOMUNICACIONES

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL
TÍTULO DE INGENIERO EN ELECTRÓNICA Y
TELECOMUNICACIONES

TEMA: DESARROLLO DE UN PROTOTIPO DE “BOT CONVERSACIONAL”
EMPLEANDO PROCESAMIENTO DE LENGUAJE NATURAL.

AUTOR: MUGLISA ROBALINO, ALEX JAVIER

DIRECTOR: ING. ALULEMA FLORES DARWIN

SANGOLQUÍ

2017

CERTIFICACIÓN

CERTIFICACIÓN



DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA
CARRERA DE ELECTRÓNICA Y TELECOMUNICACIONES

CERTIFICACIÓN

Certifico que el trabajo de titulación, "DESARROLLO DE UN PROTOTIPO DE BOT CONVERSACIONAL EMPLEANDO PROCESAMIENTO DE LENGUAJE NATURAL" ha sido revisado en su totalidad y analizado por el software anti-plagio, el mismo cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, por lo tanto, me permito acreditarlo y autorizar al señor **ALEX JAVIER MUGLISA ROBALINO** para que lo sustente públicamente.

Quito, 18 de Mayo del 2017

Atentamente,


Ing. Darwin Alulema MSc

Tutor del Proyecto de Grado

AUTORÍA DE RESPONSABILIDAD

AUTORÍA DE RESPONSABILIDAD



DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA
CARRERA DE ELECTRÓNICA Y TELECOMUNICACIONES
AUTORÍA DE RESPONSABILIDAD

Yo, **ALEX JAVIER MUGLISA ROBALINO**, con cédula de identidad N° 1722447230, declaro que este trabajo de titulación “DESARROLLO DE UN PROTOTIPO DE BOT CONVERSACIONAL EMPLEANDO PROCESAMIENTO DE LENGUAJE NATURAL” ha sido desarrollado considerando los métodos de investigación existentes, así como también se ha respetado los derechos intelectuales de terceros considerándose en las citas bibliográficas.

Consecuentemente declaro que este trabajo es de mi autoría, en virtud de ello me declaro responsable del contenido, veracidad y alcance de la investigación mencionada.

Quito, 18 de Mayo del 2017



ALEX JAVIER MUGLISA ROBALINO

NOMBRES Y APELLIDOS DEL AUTOR

C.C. 1722447230

AUTORIZACIÓN

DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA
CARRERA DE ELECTRÓNICA Y TELECOMUNICACIONES

AUTORIZACIÓN

Yo, **ALEX JAVIER MUGLISA ROBALINO**, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca Virtual de la institución el presente trabajo de titulación **“DESARROLLO DE UN PROTOTIPO DE BOT CONVERSACIONAL EMPLEANDO PROCESAMIENTO DE LENGUAJE NATURAL”** cuyo contenido, ideas y criterios son de mi autoría y responsabilidad.

Quito, 18 de Mayo del 2017



ALEX JAVIER MUGLISA ROBALINO

NOMBRES Y APELLIDOS DEL AUTOR

C.C. 1722447230

DEDICATORIA

Este gran logro es dedicado a mis padres Patricio y Mariana, a mi hermano Andrés, de igual manera a mi esposa Jacqueline e hija Dominique que gracias a ellos por medio de su apoyo y motivación he logrado cumplir una meta muy importante para mi vida profesional
Alex Javier Muglisa Robalino

AGRADECIMIENTO

Primeramente, agradezco a Dios por ser mi luz y mi camino, a mis padres por ser mi guía de todos mis pasos que he realizado y los que tengo por realizar siendo un gran apoyo para el cumplimiento de mis metas, a mi hermano y esposa por estar siempre estar al pendiente de mí.

Agradezco también al Ing. Darwin Alulema director de mi tesis, por ser mi guía y lograr el cumplimiento de este proyecto muy significativo para mi vida.

A toda mi familia y amigos porque siempre me han brindado su confianza incondicional.

Alex Javier Muglisa Robalino

ÍNDICE

CERTIFICACIÓN	ii
AUTORÍA DE RESPONSABILIDAD	iii
AUTORIZACIÓN	iv
DEDICATORIA	v
AGRADECIMIENTO	vi
ÍNDICE	vii
ÍNDICE DE TABLAS	x
ÍNDICE DE FIGURAS	xii
RESUMEN	xiv
ABSTRACT	xv
CAPÍTULO 1	1
Antecedentes	1
Justificación e Importancia	2
Alcance	3
Objetivos	4
1.1.1. General	4
1.1.2. Específicos	4
Estado del Arte	5
CAPÍTULO 2	7
Introducción	7
Funcionalidad de los bots conversacionales	7
1.1.3. Clasificación de los bots conversacionales	8
1.1.4. Características principales de los bots conversacionales	9
1.1.5. Estructura general de los bots conversacionales	9

1.1.6. Conflictos en la interacción con el usuario	10
1.1.7. Áreas de desarrollo de bots conversacionales	10
Algoritmos para PLN en bots conversacionales.....	11
1.1.8. Procesamiento del Lenguaje Natural	11
1.1.9. Modelo Probabilístico	12
Herramientas para el desarrollo del PLN.....	13
1.1.10. PHP	14
1.1.11. MySQL.....	19
1.1.12. XAMPP	20
1.1.13. HTML	21
1.1.14. JavaScript	23
Resumen del Software a utilizar	25
CAPÍTULO 3	26
Descripción general del proyecto	26
Diseño del proyecto	27
1.1.15. Descripción del algoritmo general	27
Descripción de requisitos.....	37
Diseño de interfaz web	38
Consideraciones de diseño	38
1.1.16. Diseño del Filtro para tabla profesores	38
1.1.17. Uso del índice de Jaccard para el proceso de filtrado	40
1.1.18. Métodos de ingreso de texto	44
1.1.19. Manejo de caracteres especiales.....	45
CAPÍTULO 4	46
Introducción.....	46
Escenario de prueba	46

1.1.20. Escenario 1: Determinación de la eficacia del algoritmo.....	46
1.1.21. Escenario de pruebas 2: Pruebas de servidor	58
1.1.22. Escenario de Prueba 3: Visualización	62
1.1.23. Escenario de prueba 4: Usabilidad.....	65
CAPÍTULO 5	72
5.1 Conclusiones	72
5.2 Recomendaciones.....	73
5.3 Trabajos futuros.....	73
REFERENCIAS	75

ÍNDICE DE TABLAS

Tabla 1. Clasificación de los bots conversacionales.	8
Tabla 2. Características de un bot conversacional.	9
Tabla 3. Conflictos de interacción.	10
Tabla 4. Áreas de desarrollo.	11
Tabla 5. Modelos de recuperación de información.	12
Tabla 6. Comparación de lenguajes de programación.	14
Tabla 7. Características de PHP.	16
Tabla 8. Ventajas de PHP.	17
Tabla 9. Capas de análisis de Nlp.	18
Tabla 10. API de Nlp.	18
Tabla 11. Ventajas de MySQL.	19
Tabla 12. Estándares de HTML.	21
Tabla 13. Sintaxis de programación.	24
Tabla 14. Detalle de Software.	25
Tabla 15. Tablas de la base de datos.	29
Tabla 16. Parámetros de comparación de información.	35
Tabla 17. Descripción de requisitos.	37
Tabla 18. Diseño de filtros para la tabla profesores.	39
Tabla 19. Uso del índice de Jaccard en el P.F.	40
Tabla 20. Comparación de métodos de ingreso de texto.	44
Tabla 21. Conflicto de caracteres especiales en diferentes lenguajes.	45
Tabla 22. Descripción de las pruebas para la tabla Profesores.	46
Tabla 23. Resultados de Prueba 1.	47
Tabla 24. Resultados de Prueba 2.	48
Tabla 25. Resultados de Prueba 3.	48
Tabla 26. Eficiencia del algoritmo para la tabla Profesores.	49
Tabla 27. Descripción de las pruebas para la tabla Materias.	50
Tabla 28. Resultados de Prueba 1.	50
Tabla 29. Resultados de Prueba 2.	50
Tabla 30. Eficiencia del algoritmo para la tabla Materias.	52

Tabla 31. Descripción de las pruebas para la tabla Ubicación.....	52
Tabla 32. Resultados de Prueba 1.	52
Tabla 33. Eficiencia del algoritmo para la tabla Ubicación.	53
Tabla 34. Descripción de las pruebas para la tabla Eventos.	54
Tabla 35. Resultados de Prueba 1.	54
Tabla 36. Eficiencia del algoritmo para la tabla Eventos.....	55
Tabla 37. Descripción de las pruebas para la tabla Eventos.	56
Tabla 38. Resultados de Prueba 1.	56
Tabla 39. Eficiencia del algoritmo para la tabla Graduación.....	57
Tabla 40. Eficiencia total del algoritmo.	58
Tabla 41. Tipos de prueba en el servidor local.	58
Tabla 42. Hosting para el servidor externo.	60
Tabla 43. Características de los dispositivos empleados.	62

ÍNDICE DE FIGURAS

Figura 1. Estructura general de un bot conversacional.	9
Figura 2. Índice de Jaccard.....	13
Figura 3. Logotipo de PHP.	15
Figura 4. Estructura de ejecución de Scripts.....	15
Figura 5. Estructura de MySQL.	20
Figura 6. Estructura de Programación.....	23
Figura 7. Esquema General del aplicativo.	26
Figura 8. Diagrama de casos de la aplicación web.	27
Figura 9. Descripción del algoritmo general.....	28
Figura 10. Descripción de la tabla profesor.	30
Figura 11. Descripción de la tabla ubicación.	30
Figura 12. Descripción de la tabla graduación.....	30
Figura 13. Descripción de la tabla eventos.	30
Figura 14. Descripción de la tabla materias.	31
Figura 15. Código de conexión a base de datos.	31
Figura 16. Proceso de conexión con base de datos.	32
Figura 17. Función de tokenización.	32
Figura 18. Proceso de tokenización.	33
Figura 19. Función de cálculo de índice de Jaccard.....	34
Figura 20. Proceso de clasificación.....	34
Figura 21. Proceso de comparación de información.....	36
Figura 22. Proceso de impresión.....	37
Figura 23. Esquema de la página web.....	38
Figura 24. Anidación de Jaccard para tabla profesores.....	42
Figura 25 . Anidación de Jaccard para tabla ubicación.....	42
Figura 26. Anidación de Jaccard para tabla graduación.	43
Figura 27. Anidación de Jaccard para tabla eventos.	43
Figura 28. Anidación de Jaccard para tabla materias.....	44
Figura 29. Resultados tabla Profesores.	48

Figura 30. Promedio de respuesta para tabla Profesores.	49
Figura 31. Resultados Tabla Materias.....	51
Figura 32. Promedio de respuesta Tabla Materias.	51
Figura 33. Resultados tabla Ubicación.....	53
Figura 34. Promedio de respuesta para tabla Ubicación	53
Figura 35. Resultados tabla Eventos	55
Figura 36. Promedio de respuesta para tabla Eventos.	55
Figura 37. Resultados tabla Graduación.	57
Figura 38. Promedio de respuesta para la tabla Graduación.....	57
Figura 39. Estructura del escenario de pruebas del servidor local.....	59
Figura 40. Ejecución de la aplicación en LAN.	59
Figura 41. Historial del chat.....	60
Figura 42. Estructura del escenario de pruebas del servidor externo.....	61
Figura 43. Ejecución del aplicativo en el servidor externo.....	61
Figura 44. Respuesta de consulta en servidor externo.	62
Figura 45. Visualización en computador1 en servidor local.....	63
Figura 46. Visualización en computador2 en servidor local.....	63
Figura 47. Visualización en tablet en servidor local.....	63
Figura 48. Visualización en Smartphone en servidor local.....	64
Figura 49. Visualización en computador 1 en servidor web.....	64
Figura 50. Visualización en computador 2 en servidor web.....	64
Figura 51. Visualización en tablet en servidor web.	64
Figura 52. Visualización en Smartphone en servidor web.....	65
Figura 53. Resultado de la pregunta 1.....	66
Figura 54. Resultado de la pregunta 2.....	66
Figura 55. Resultado de la pregunta 4.....	67
Figura 56. Resultado de la pregunta 6.....	67
Figura 57. Resultado de la pregunta 3.....	68
Figura 58. Resultado de la pregunta 5.....	68
Figura 59. Reajuste de la tabla profesores.	69
Figura 60. Reajuste de la tabla materias.....	70
Figura 61. Resultado de la pregunta 7.....	70

RESUMEN

El uso de la Inteligencia Artificial en los últimos tiempos se ha convertido en un foco de desarrollo y generación de información, tal es el caso de que, el uso de bots conversaciones están actualmente suprimiendo al personal dedicado en los sectores que brindan atención al cliente como es el caso: de las entidades bancarias, puntos de información, ventas y traductores de idiomas; también se han desarrollado aplicaciones que brindan respuestas automáticas para las principales redes sociales como asistentes personales. Este proyecto brinda información a manera de chat en tiempo real mediante el uso de librerías de Procesamiento del Lenguaje Natural para PHP, donde las consultas realizadas por el usuario en un programa de servidor web (local y externo) pasan por un proceso de filtrado donde son Tokenizadas para posteriormente calcular el Índice de Jaccard, para permitir el acceso una base de datos en el cual se almacena la información relativa, en los casos de ambigüedad en la pregunta, este proceso se repite con las características principales de la búsqueda, que automáticamente se presentan como respuesta en un campo de texto dentro de la aplicación web. Este trabajo concede el acceso a la información de una forma ordenada y automática facilitando al Departamento de Electrónica y Electrónica de la Universidad de las Fuerzas Armadas ESPE los tiempos de respuesta a inquietudes por parte de los estudiantes u otros usuarios.

PALABRAS CLAVE:

- *INTELIGENCIA ARTIFICIAL*
- *PROCESAMIENTO DEL LENGUAJE NATURAL*
- *CHATBOT*
- *NLTK*
- *ÍNDICE DE JACCARD*

ABSTRACT

The use of Artificial Intelligence in recent times has become a focus of development and generation of information, such is the case that the use of bots conversations are currently suppressing dedicated staff in the sectors that provide customer service as Is the case: from banks, information points, sales and language translators; applications have also been developed that provide automatic responses to the main social networks as personal assistants. This project provides real-time chat information through the use of Natural Language Processing libraries for PHP, where queries by the user in a web server program (local and external) go through a filtering process where they are In order to calculate the Jaccard index, to allow access to a database in which the relative information is stored, in cases of ambiguity in the question, this process is repeated with the main characteristics of the search, which automatically They present as a response in a text field within the web application. This work grants access to the information in an orderly and automatic way, providing the Department of Electronics and Electronics of the Universidad de las Fuerzas Armadas – ESPE the times of response to concerns by students or other users.

KEYWORDS:

- **ARTIFICIAL INTELLIGENCE**
- **PROCESSING OF NATURAL LANGUAGE**
- ***CHATBOT***
- ***NLTK***
- **JACCARD INDEX**

CAPÍTULO 1

Antecedentes

A principios del año 1950, aparecen los primeros programas computacionales basados en inteligencia artificial, algunas de estas aplicaciones centradas en la creación de máquinas virtuales capaces de interpretar e interactuar mediante Lenguaje Natural (Cobos Torres, 2013). Alan Turing propuso en el año de 1956, evaluar la conversación entre un humano y una máquina para determinar el grado de certeza a una pregunta formulada por una tercera persona, mediante los análisis de realización durante 5 minutos de conversación textual se verificó que dicha máquina presentaba respuestas similares a las que daría un ser humano (Molina & Torres, 2015).

La inteligencia artificial en el transcurso del tiempo ha venido evolucionando a tal punto que en la actualidad es aplicada en distintas áreas como en medicina, ingeniería, entretenimiento, entre otros, con el fin de facilitar el día a día de las personas. El Procesamiento del Lenguaje Natural (PLN) une las tecnologías de la ciencia computacional como la inteligencia artificial, la lingüística aplicada; con el propósito de hacer posible la comprensión y el procesamiento asistido por el ordenador expresada en lenguaje humano para determinadas tareas (Rodríguez & Benavidez, 2013).

En la actualidad existen varias aplicaciones de procesamiento de Lenguaje Natural en el ámbito Tecnológico-social; entre asistentes virtuales personales más conocidos están, Siri prototipo realizado en iOS y Google Now para la plataforma perteneciente a Android, con la funcionalidad de contestar preguntas o hacer recomendaciones sobre algún argumento dado por el usuario (Jimenez & Sanchez, 2015).

Justificación e Importancia

De acuerdo al Plan Nacional del Buen Vivir 2013-2017, referente al objetivo 11, de los mercados de telecomunicaciones, detalla que la mayoría de las instituciones públicas y privadas en el país no cuentan con servicios electrónicos, es por eso la importancia de desarrollar aplicaciones de inclusión social, pero es el caso de las personas de tercera edad o discapacitados tengan la facilidad de realizar consultas electrónicas sin la necesidad de acudir a un punto de información (Secretaría del Buen Vivir, 2013).

En América del Sur, al 2012, Ecuador se encontraba en el puesto ocho de doce países considerados para el análisis del nivel de desarrollo de gobiernos electrónicos. Este resultado evidencia el nulo avance que el país ha presentado en cuanto al desarrollo de este tipo de tecnologías, que lo ha hecho descender siete lugares en los dos últimos años. Pretendiendo bajo el objetivo 11.3 sección i anuncia: “impulsar el gobierno electrónico transaccional participativo para que la ciudadanía acceda en línea a datos, información, trámites y demás” (Secretaría del Buen Vivir, 2013).

El Lenguaje Natural juega un papel importante para la creación de diferentes aplicativos, como el desarrollo de bots conversacionales, siendo una rama de la inteligencia artificial, basada en sistemas informáticos, creando una comunicación entre el hombre y un computador (texto y voz) por medio del lenguaje humano, que a diferencia del lenguaje de máquina el cuál emplea diferentes números de códigos interpretados con lo que se lograría la interacción entre las personas y sistemas computacionales de una forma más natural (Torres Moreno, 2016).

Existen varios tipos de algoritmos empleados para el procesamiento de Lenguaje Natural, uno de ellos es el algoritmo Booleano fundamentado en que criterio de decisión binaria, que básicamente consiste en definir dicha información como un grupo de términos de indexación o *keywords*, otro de los algoritmos empleados es de forma vectorial que está basado en la construcción de una matriz

de palabras y documentos, en la cual las filas reflejan los documentos almacenados en una base de datos mientras que en las columnas de la matriz poseen los términos adecuados para cada documento y por ultimo existe el algoritmo probabilístico , que se basa en el cálculo de probabilidad de un documento sea importante al formularse algún interrogante tareas (Rodriguez & Benavidez, 2013).

Hay algunas herramientas que por a su funcionalidad permiten el desarrollo de aplicaciones basadas en el procesamiento de Lenguaje Natural, como Python que contiene una plataforma que es NLTK (*Natural Language Tool kit*) que provee interfaces como Word Net (recurso léxico) y otras que obvian el uso de bases de datos para almacenamiento de palabras, pero con las desventajas de dificultad de creación de interfaces gráficas y la no portabilidad del software (Drake, 2016).

Alcance

Se analiza las características y funcionalidades de diferentes algoritmos que son aplicados al procesamiento de Lenguaje Natural, con el propósito de determinar el más óptimo y fácil de implementar para el diseño del bot conversacional en la web para que al momento de realizar cualquier tipo de pregunta no exista problemas de ambigüedad por el contexto del mensaje.

De igual manera existen varias herramientas de software que se indaga con la finalidad de elegir la más apropiada para el aplicativo del bot conversacional procurando que no existan inconvenientes como el uso del lenguaje en español ya que en la gran mayoría usan librerías en inglés.

Para el diseño del prototipo del bot conversacional, se considera almacenar la información, además se diseñar una interfaz web que sea interactiva y fácil de usar. Con relación a la prueba de concepto, se validará mediante el diseño de un aplicativo, que permita a los usuarios acceder al aplicativo vía browser con el propósito de interactuar con el bot conversacional

Esta aplicación se realiza con información del Departamento de Eléctrica y Electrónica de la Universidad de las Fuerzas Armadas - ESPE, que tiene como finalidad responder diferentes interrogantes, como materias que pertenecen a cada nivel de las carreras, horarios de clases, información, eventos; entre otros aspectos, mediante el empleo del Procesamiento de Lenguaje Natural, obteniendo la ventaja de no acudir directamente a las instalaciones procurando mejorar el tiempo de respuesta a las inquietudes de los estudiantes y público en general.

Objetivos

1.1.1. General

- Desarrollar un prototipo de "Bot Conversacional" para consultas informativas sobre datos y entidades en el departamento de Eléctrica y Electrónica de la Universidad de las Fuerzas Armadas empleando Procesamiento de Lenguaje Natural.

1.1.2. Específicos

- Definir características de Bots conversacionales.
- Determinar el algoritmo de Procesamiento de Lenguaje Natural para la implementación del bot conversacional.
- Definir las herramientas de programación para el desarrollo de un bot conversacional basado en el Procesamiento del Lenguaje Natural.
- Definir características para la prueba de concepto del bot conversacional.
- Configurar una base de datos para el almacenamiento de información mediante la interacción con los usuarios y datos de la prueba de concepto.
- Configurar el servidor web para ejecutar el aplicativo diseñado.

Estado del Arte

En los años 50 Alan Turing en (Turing, 1950) propuso “*The Imitation Game*” que consistió en encontrar la respuesta a la pregunta ¿las máquinas pueden pensar?, mediante una serie de juegos dentro de un interrogatorio. En 1969 en los laboratorios del MIT (*Massachusetts Institute of Technology*) se desarrolló el primer bot conversacional como simulación de un psicoterapeuta (Shrager, 2015), continuó con su implementación hasta el año 1979, contribuyendo a la inteligencia en el diseño de varios juegos de computadoras en décadas posteriores.

El SRI (*Stanford Research Institute*) para el año 2008 emprende “Siri Project” que es un asistente personal para Smartphone basado en la inteligencia artificial para la comunicación con el usuario, considerado como una de las más grandes innovaciones tecnológicas (Winarsky, Mark, & Kressel, 2013) ya que podía convertir el sonido de la voz a texto para entender y responder los requerimientos del usuario (President's Council of Advisors on Science and Technology, 2010).

En la actualidad existen varios bots conversaciones desarrollados para diferentes aplicaciones y con distintas capacidades de respuesta dependiendo del diseño realizado por el programador (Ramírez, Rubén;, 2011). A nivel de Smartphone se han implementado asistentes personales los cuáles en un principio estaban enfocados a reemplazar las herramientas de búsqueda dentro del dispositivo combinando la interacción del teclado y voz siendo una evolución de los chatterbots (Bykarskij, 2015).

Empresas como Google, Microsoft y Apple han extendido este tipo de asistentes en función de cada una de las necesidades dentro de sus sistemas, como son: Google Now, Cortana y Siri respectivamente; permitiendo así que el usuario interactúe en distintos idiomas en base a la activación y reconocimiento de voz por medio del uso de herramientas de Procesamiento del Lenguaje Natural (Hunter, 2015; Jiménez & Sánchez , 2015), esto ha permitido que este tipo ayudantes virtuales den el salto a sistemas operativos de escritorio.

Durante el año 2016 se presentaron varios proyectos de domótica como es Google “Home” el cuál integra al asistente Google Now siendo el cerebro para la automatización de una casa, diseño que se encuentra actualmente en desarrollo (Bohn, 2016); pero Amazon ya presentó en año 2015 a “Echo” con su servicio de voz llamada Alexa, la que permite la comunicación y realización de cierto tipo de peticiones por el usuario como la acceder a aplicaciones de streaming musical y brindar información como noticias, resultados deportivos, clima entre otros (Amazon Inc, 2015; Quentin, 2015).

La compañía JiboInc se encuentra implementado al primer robot social del mundo, amigable de utilidad y con inteligencia; que puede: ver, mediante el uso de cámaras aplicadas a técnicas de reconocimiento facial y captura de imágenes; escuchar y hablar, por medio el uso de micrófonos, parlantes y algoritmos de Procesamiento del Lenguaje Natural; aprender, se adapta a través de la inteligencia artificial a las acciones cotidianas del usuario (Jibo Inc , 2016). Actualmente la empresa actualizó su SDK, la cual incluye avances en las características mencionadas anteriormente, brindando al robot destrezas creativas que mejoran la interacción con distintos usuarios; las próximas actualizaciones prevén incluir un API que complemente las capacidades de video y audio de Jibo, además el reconocimiento del lenguaje, de voz, conversiones de texto a audio, geolocalización y corrección de errores en la conectividad web (Jibo Inc, 2016).

Estos dispositivos actualmente se encuentran en proceso de implementación para el idioma Inglés y no existen registros para su expansión a otros idiomas como el Español.

CAPÍTULO 2

Introducción

La inteligencia artificial está vinculada con el Procesamiento del Lenguaje Natural siendo una de las principales áreas la investigación y desarrollo de bots conversacionales principalmente en los ambientes educativos, sociales y administrativos (García & Valle, 2013) en tiempo real permitiendo al usuario tener respuestas inmediatas y coherentes. Los recientes avances tecnológicos se han desarrollado para distintas aplicaciones que usan el Procesamiento del Lenguaje Natural como búsquedas de documentos requeridos hasta robots sociales que respondan a inquietudes formuladas por el usuario, o tan solo establezcan una conversación (Cortez, Vega, & Pariona, 2013).

Por medio del test de inteligencia realizado por Alan Turing (Turing, 1950) , se promovió en el transcurso de los años el desarrollo de bots conversacionales con más parámetros de diseño; siendo prácticos y adaptables a la vida cotidiana del usuario, dependiendo del entorno en el que este se encuentre.

Actualmente, existe una gran variedad de herramientas de programación que permiten el diseño de bots conversacionales, como: AIML (Artificial Intelligence Mark-up Language), Python, Java, Ruby, PHP, entre otros (PHP, 2016). Siendo los lenguajes propicios para el desarrollo web con concatenaciones a MySQL; encargada de la gestión de almacenamiento de información para el Procesamiento del Lenguaje Natural.

Funcionalidad de los bots conversacionales

Tecnológicamente los bots conversacionales son considerados programas automáticos que simulan una conversación “online” utilizando el Lenguaje Natural de las personas mediante el uso de algoritmos, diccionarios lingüísticos, combinación de técnicas y teorías sobre el Procesamiento del Lenguaje Natural, actualmente se está desarrollando la tecnología multimedia cuyo objetivo es añadir a todas estas

características nombradas previamente, estadísticas de animación, análisis de gráficos y sonidos captados durante la interacción del usuario y la máquina (Aref, 2005).

1.1.3. Clasificación de los bots conversacionales

Sus principales funcionalidades son: responder preguntas o realizar acciones (Aetecno, 2016), se los puede clasificar en función del área de servicio, como se describe en la tabla 1 (Cobos, 2013; Neoteo, 2012):

Tabla 1.
Clasificación de los bots conversacionales.

Área	Descripción
Social	En la actualidad se ha desarrollado el aspecto social de los bots como asistentes virtuales brindando atención al cliente.
Académica	Responde preguntas relacionadas a temas educativos, al igual que permite traducciones automáticas en diferentes idiomas.
Ocio	Principalmente usado para la diversión como juegos y manejo de redes sociales.
Medicina	Enfocado en ser un asistente virtual permitiendo realizar consultas y generando recetas para diferentes enfermedades “generales”.
Empresarial	Diseñado para el estudio de mercado, ventas y generación de campañas publicitarias.

1.1.4. Características principales de los bots conversacionales

Las principales características de los bots conversacionales se detallan en la tabla 2 (Cobos, 2013; Vela, 2015):

Tabla 2.
Características de un bot conversacional.

Característica	Descripción
Adaptabilidad	Capacidad de cambiar comportamiento en base al aprendizaje.
Personalidad	Diferentes características del comportamiento verbal o escrito dependiendo del programador.
Racionalidad	Realiza funciones correctas a partir de los datos recibidos.
Veracidad	No proporciona información falsa.

1.1.5. Estructura general de los bots conversacionales

La estructura general de los bots conversacionales se detalla en la figura 1.

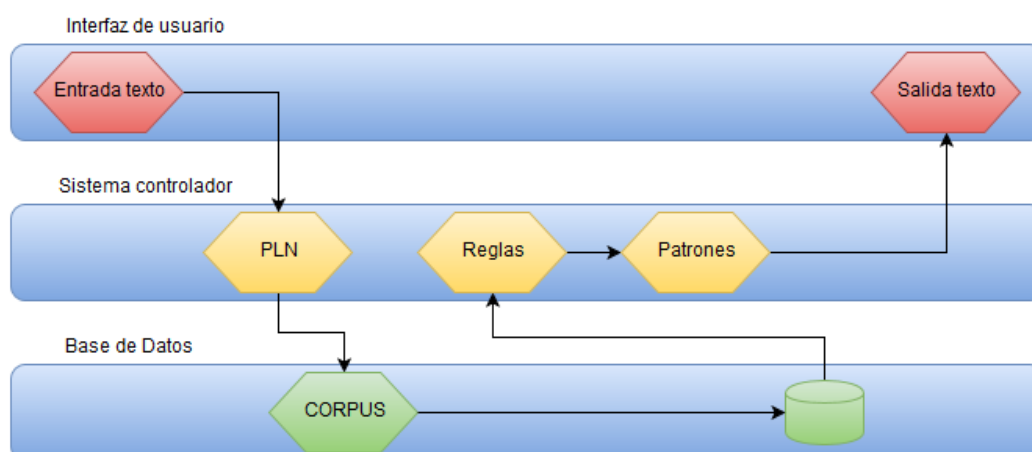


Figura 1. Estructura general de un bot conversacional.

El texto ingresa mediante una interfaz gráfica de usuario y es procesado en un sistema de control baso en Procesamiento del Lenguaje Natural que, mediante una

conexión a la base de datos, compara el corpus con la información almacenada y está en función de reglas preestablecidas fija patrones para generar una respuesta y presentar al usuario.

1.1.6. Conflictos en la interacción con el usuario

Al ser los bots conversacionales desarrollados para un propósito en concreto, muchas veces pueden estos asumir conceptos erróneos cuando el usuario interactúa de una forma diferente a las reglas o patrones preestablecidas por el programador (Polatidis, 2011), Los principales problemas se detallan en la tabla 3:

Tabla 3.
Conflictos de interacción.

Inconveniente	Detalle
Sobre carga de preguntas	Se presenta cuando la base de datos no soporta el número de preguntas consecutivas realizadas por el usuario.
Ambigüedad	Existe cuando el bot no puede distinguir el contexto de una pregunta formulada.
Errores gramaticales	Concurre cuando el texto posee errores de puntuación.
Ortografía	Mala escritura realizado por el usuario.
Uso de anáforas	Uso repetitivo de términos dentro de las preguntas.
Lenguaje coloquial	Se presenta cuando el programador no incluye términos locales de una región.
Anglicismo	Importación de términos del idioma Inglés al Español.

1.1.7. Áreas de desarrollo de bots conversacionales

En los últimos años la inteligencia artificial se está desarrollando en ámbitos tecnológicos, comerciales, científicos e investigativos; con finalidades de diseñar aplicaciones centradas en agentes conversacionales para diversos ambientes con el

propósito de facilitar el diario vivir al ser humano, estas áreas son detalladas en la tabla 4 (Russell & Norvig, 2004; Pazos & Barreiro, 1999):

Tabla 4.
Áreas de desarrollo.

Característica	Descripción
Agente solucionador de problemas	Toma decisiones en función de patrones preestablecidos o programados por medio de algoritmos diseñados.
Aprendizaje Automático	Mediante el uso de algoritmos de procesamiento de información el bot es capaz de aprender datos del usuario y generar respuestas en función de las costumbres más comunes.
Entorno de funcionalidad	Por medio del uso de máquinas de estado y árboles de decisión genera respuestas y acciones a las situaciones planteadas u originadas.
Minería de información	Realiza búsquedas en otras fuentes para presentar la mejor respuesta, debido a que la información consultada no se encuentra dentro de su base de datos.
Manejo de sistemas	Realiza las tareas de operatividad dentro de un sistema mecatrónico, operativo, domótica, entre otros.

Algoritmos para PLN en bots conversacionales

1.1.8. Procesamiento del Lenguaje Natural

Como una disciplina de la inteligencia artificial combinada con la lingüística, empezó su investigación en la década de los 60 el Procesamiento del Lenguaje Natural con la finalidad de estudiar la problemática que es derivada de la generación del

Lenguaje Natural, es decir, buscar la comprensión del habla en lenguaje materno de cada individuo (Vallez, 2007; Mateos & Reina, 2013).

En los últimos años con el desarrollo tecnológico de la informática, esto se modernizó a buscar la interrelación entre el humano y la máquina. Los agentes conversacionales se encuentran estrictamente ligados con las tecnologías de Procesamiento del Lenguaje Natural basándose en algoritmos de extracción de información, toma de decisiones y aprendizaje (Van Lun, 2014).

1.1.8.1. Modelos de recuperación de información

Existen 3 modelos para la recuperación de la información, detallados en la tabla 5 (Russell & Norvig, 2004; Blázquez, 2012):

Tabla 5.
Modelos de recuperación de información.

Modelo	Característica
Probabilístico	Mediante el cálculo de probabilidades busca determinar la importancia de un texto para la consulta realizada.
Vectorial	Representa al documento en forma de vector con la finalidad de identificar las consultas y la relevancia de la misma, se puede ampliar a varios documentos mediante el uso de matriz que puede considerarse una base de datos.
Binario	Crea grupos de palabras y mediante la toma de decisiones busca determinar si existe en el texto la consulta realizada

1.1.9. Modelo Probabilístico

Para aplicar este modelo se necesita crear un conjunto de términos y cada uno de estos debe tener un valor independiente con relación al área de búsqueda, utilizando la

formula básica de la probabilidad $P= n/N$; en la cual se divide una parte para el todo; este valor se encuentra entre 0 y 1 donde 1 es la presencia total en todo el documento y 0 representa la ausencia ; n es cada término del conjunto previamente determinado para la búsqueda, y N es el documento completo donde se realiza la indagación (Blázquez, 2012).

1.1.9.1. Índice de Jaccard

Este índice relaciona el número de elementos de un conjunto compartidas con el número total de elementos exclusivos, diseñado para ser igual a uno es casos de similaridad completa e igual a cero sino existen elementos en común; es decir la intersección de dos conjuntos sobre la unión de los mismos, proceso que se detalla en la figura 4 (Chao, 2004).

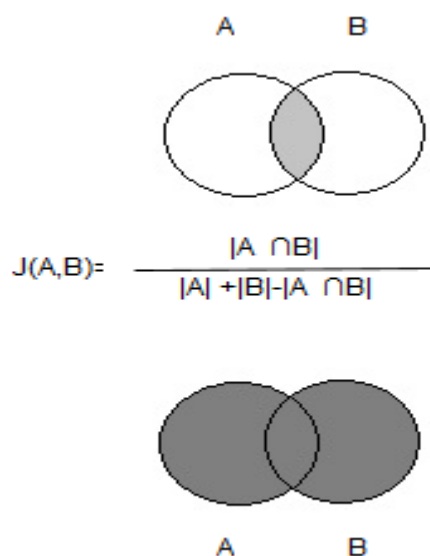


Figura 2. Índice de Jaccard.

Herramientas para el desarrollo del PLN

El análisis de PLN se puede realizar en sin número de lenguaje de programación, donde cada uno brinda ciertos beneficios con relación a otros: esto debido a interés de

los desarrolladores en la creación de APIs y librerías; en la tabla 6 se presenta una comparativa entre Python, Java y PHP.

Tabla 6.
Comparación de lenguajes de programación.

Lenguaje	Librerías desarrolladas	Características
Python	<ul style="list-style-type: none"> • Nltk • SpaCy 	Se requiere la creación de un servidor dedicado solo para el PLN, debido a las limitaciones que este lenguaje posee dentro del diseño de páginas web.
Java	<ul style="list-style-type: none"> • OpenNLP • StandfordNLP 	Las librerías desarrolladas dependen de una gran importación de clases y alto conocimiento del lenguaje para poder ejecutar las herramientas de PLN.
PHP	<ul style="list-style-type: none"> • Nlptools • PHPStandford 	Debido a su gran integralidad y extensa documentación para el desarrollo de sistemas web, permite realizar aplicaciones de trabajo en tiempo real.

1.1.10. PHP

Sus siglas significan “*Hypertext Preprocessor*”. Nace a mediados de los 90 como un lenguaje de programación web que se ejecuta en un servidor, con capacidad de comunicación con base de datos para el desarrollo de aplicaciones sencillas, permitiendo así páginas web dinámicas (Peinado, 2010).



Figura 3. Logotipo de PHP.
Fuente: The PHP Group.

Para el año 2015 se lanzó la versión 7.0 de este lenguaje el cual incluye mejoras en el tipo de declaración de funciones, pero no es la más difundida, ya que la versión 5.0 posee actualmente un soporte universal en todo tipo de servidores, APIs, librerías y documentación, siendo la más óptima para cualquier tipo de desarrollo web (González, 2016).

1.1.10.1. Características, ventajas, desventajas y estructura de programación

La estructura de ejecución de un Script de PHP se detalla en la figura 4, el usuario realiza una petición al servidor mediante el uso de formularios HTML, que se conectan con el código PHP, donde generalmente son consultas a la información de la base de datos, estos son procesados como la respuesta al formulario en función de los requerimientos del mismo y son presentados en una nueva página HTML.

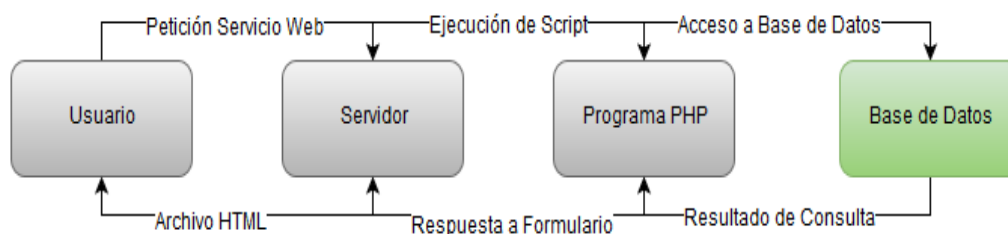


Figura 4. Estructura de ejecución de Scripts.

La estructura de programación en PHP, es simple, requiere de una etiqueta inicial y final para delimitar el código (<?, para iniciar el código; ?>, para finalizar el código). Para la realización de páginas web más dinámicas se unen los lenguajes de programación HTML y PHP, siendo este último embebido con conexiones a

formularios (botones, casillas de verificación y entre otros) dentro de HTML a manera de Script, contribuyendo a una rápida visualización web (Vílchez, 2007).

PHP diferencia dentro de la declaración de variables: caracteres alfanuméricos, símbolos y diferencia entre mayúsculas y minúsculas (Vázquez, 2008). Los tipos de datos soportados por este lenguaje son los siguientes (Palomo & Montero, 2011):

- Los de tipo básico: comprenden, caracteres, booleano, enteros y flotantes.
- Los de tipo compuesto: comprenden, los objetos y arreglos.
- Los tipos especiales: comprenden, nulos y variables especiales.

Las características principales de PHP, se detalla en la tabla 7 (Vázquez, 2008):

Tabla 7.
Características de PHP.

Característica	Descripción
Simplicidad	Permitir la generación de código en tiempos reducidos.
Estabilidad	Administración propia del sistema para evitar bugs en la ejecución.
Velocidad	No requiere demasiados recursos, al igual que se integra a diferentes ambientes y softwares.
Seguridad	Posee protecciones contra ataques.
Conectividad	Dispone de gran cantidad de librerías y extensiones de fácil manejo.
Documentado	Existe gran cantidad de guías de uso.
Multiplataforma	Puede ser ejecutada en cualquier sistema operativo.

Las ventajas y desventajas de programar en PHP se describen en la tabla 8 (Programando la web, 2016; Ken, 2015; Santa María, 2014; Paz, 2012):

Tabla 8.
Ventajas de PHP.

Ventajas	Desventajas
Lenguaje totalmente abierto y libre.	Necesidad de un servidor propio.
Multiplataforma.	Conexión a base de datos no está estandarizada.
Alto soporte.	Código fuente no puede ser ocultado.
Dinámico.	
Amplia documentación.	
Fácil acceso a base de datos.	
Múltiples extensiones y librerías.	
Rápido y seguro.	

1.1.10.2. *Php-nlp-tools*

Es una herramienta de PHP encargada del Procesamiento del Leguaje Natural, que provee elementos básicos para la formación de modelos mediante una organización; por medio del archivo “composer.json”, siendo un operador de dependencias que asocia diferentes librerías pertenecientes a otras herramientas, para garantizar el correcto funcionamiento del aplicativo, generalmente es instalado en el directorio vendor (PHP, 2015; Styde, 2014). Al análisis que realiza Nlp, se lo realiza por capas, detalladas en la tabla 9 (Buitelaar, 2014):

Tabla 9.
Capas de análisis de Nlp.

Capa	Descripción
Part of Speech	Encargada de catalogar, etiquetar cada estructura del texto.
Etiqueta semántica	Se relaciona con la búsqueda del significado lingüístico de cada estructura.
Funciones gramaticales	Basada en reglas de la gramática, realiza un análisis y busca errores.

Dentro de la herramienta de php-nlp-tools existen varios API (*Application Programming Interface*), detallados en la tabla 10 (PHP, 2015):

Tabla 10.
API de Nlp.

API de Nlp	Descripción
Análisis	Existen dos funciones: Distribución de frecuencia para obtención de datos e identificador que aumenta el nivel de frecuencia.
Clasificadores	Hay dos funciones: Clasificador lineal, clasifica información de un modelo lineal y Clasificador multinomial para la clasificación de documentos.
Tokenización	La función es basada en un clasificador de tokenizador que describe si existe el fin de una palabra y expresiones reguladas basada en un conjunto de texto y espacios en blanco tokenizador.
Utilidades	Trasformación basada en un clasificador, encargado de aprobar y pasar un grupo de datos basados en una clase, palabras vacías no contienen ninguna información y grupo de vocales abstractas.

1.1.11. MySQL

Es un código abierto que permite la gestión de base de datos, disponible para múltiples plataformas. Fue diseñado con la intención de conectar tablas mediante el uso de rutinas rápidas de bajo nivel, considerado en ese tiempo como una necesidad flexible y rápida, desarrollada en la década de los 90 por Michael Widenis (Oracle, 2011).

1.1.11.1. Características, ventajas, desventajas y estructura de programación

Ha ganado gran popularidad debido a sus grandes prestaciones como son (Oracle, 2011; Casillas, 2013):

- Administración basada en usuarios y privilegios.
- Estabilidad.
- Sistema de reserva de memoria basado en hilos.
- Gran portabilidad.
- Velocidad y robustez.
- Multiplataforma.
- Dispone de APIs para un sin número de lenguajes de programación.
- El servidor puede estar instalado en cualquier equipo.
- Amplio soporte al idioma Español, dentro de las sentencias.

Las principales ventajas y desventajas de utilizar MySQL, se detalla en la tabla 11 (Enríquez & Maldonado, 2010):

Tabla 11.
Ventajas de MySQL.

Ventajas	Desventajas
Conectividad extensa con los múltiples lenguajes de programación.	No es intuitivo.
Seguridad.	Múltiple creación de usuarios, puede generar retardos en el servidor.
Poca probabilidad de corromper datos.	Limitaciones de sentencias.
Fácil instalación y configuración.	
Alta velocidad al realizar distintas operaciones.	

Las tablas creadas de MySQL, se almacenas en una base de datos; una aplicación externa que desee acceder a esta información requiere de un Script escrito en MySQL; el cuál filtra, busca y categoriza los datos en función de la solicitud realizada accediendo a la base de datos, procese que se detallan en la figura 5 (Castillo, 2011).

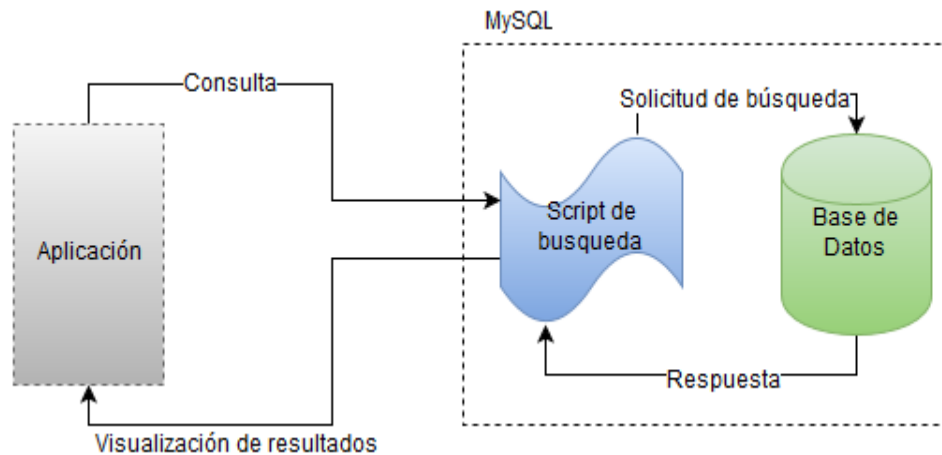


Figura 5. Estructura de MySQL.

1.1.12. XAMPP

Es un paquete independiente, multiplataforma y de software libre que permite instalar varios tipos de servidores, incluye servidor web Apache, que permite montar servicios de páginas web; servidores de bases de datos con sus respectivos gestores. Su nombre se deriva de los principales servicios que posee: X, al ser multiplataforma; A, del servidor Apache; M, de los servicios de base de datos MySQL; P, por el lenguaje de programación PHP; y P, del lenguaje Perl (Baeza, 2005).

1.1.12.1. Servidor web Apache

Se origina en el año 1995, por la NCSA (*National Center for Super Computing Applicationes*), a manera de un servidor Web elemental que corría solo sobre servidores Unix, para ser luego administrado por Apache Group; en la actualidad más del 50% de la administración Web de Internet se realiza utilizando este servicio (kabir, 2006).

La configuración se realiza mediante el uso de archivos de texto con directivas, la ubicación del archivos principal es `https.conf` el cual contiene: directivas de control, la

que permite el control del servidor; parámetros del servidor principal, determina el comportamiento del servidor en función de los parámetros requeridos; configuraciones de host virtuales, la atención de los requerimientos del servidor (Junta de Andalucía, 2010; Pavón, 2013).

Las principales características de este servidor web son (Junta de Andalucía, 2010):

- Multiplataforma.
- Diseño Modular.
- Soporta múltiples lenguajes de programación.
- Permite host virtuales.
- Fácil configuración.
- Open Source.
- Simplicidad.


1.1.13. HTML

Es un lenguaje de programación simple, dedicado para la creación de páginas web, por medio de texto y diferentes etiquetas, sus siglas significan (*HyperText Markup Language*) (MDN, 2016). Tim Berners-Lee en el año de 1989 propuso dicho lenguaje junto al protocolo HTTP con el propósito de compartir información a nivel mundial (Webadictos, 2012).

Existen cinco estándares de HTML, ya que cada una de ellas establecen normas distintas a cada etiqueta autorizada, las cuáles se detallan en la tabla 12 (Webadictos, 2012; Apr, 2016):

Tabla 12.
Estándares de HTML.

Estándar	Descripción
HTML 1	Considerada solo para pruebas, ya que no fue reconocida como estándar.

Continua 

HTML 2.0	Primer estándar de HTML, soportaba tablas y era opcional la declaración de cabecera, cuerpo y HTML.
HTML 3.2	Publicada en 1997, anexó páginas web como applets de Java.
HTML 4.0	Publicada en 1998, fue posible incluir pequeños Scripts a páginas web y se podía trabajar con tablas complejas.
HTML 5	Funciona en dos tipos de sintaxis HTML y XHTML, es posible introducir audio, video y uso de diferentes nuevas etiquetas.

1.1.13.1. Características

Las principales características del lenguaje de programación HTML se detallan a continuación (Ibiblio, 2014):

- Multiplataforma.
- Dinámico.
- No utiliza variables.
- No se compila.
- Estructura de lenguaje basado en etiquetas.

1.1.13.2. Estructura de programación.

Existen una serie de reglas sobre la estructura de programación sobre donde pueden ubicarse o no los elementos; es decir ciertos elementos no pueden contener a otros ni ser solapados (Web Docente Departamental, 2010), como se detalla en la figura 6.

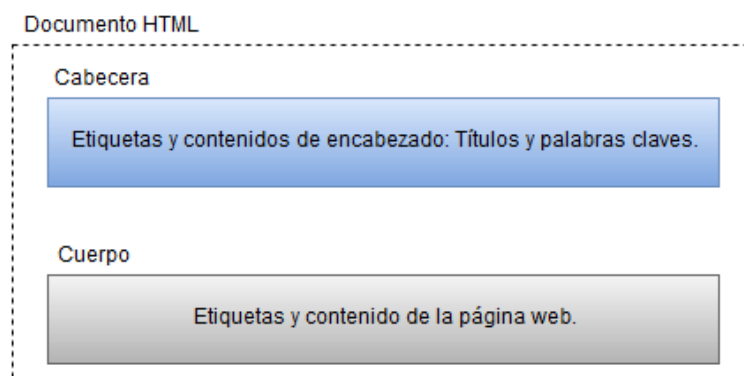


Figura 6. Estructura de Programación.

1.1.14. JavaScript

Fue implementado por Netscape en los años 90, basado en Java, su principal aplicación es el diseño de páginas web, pero este solo funciona dentro de una página HTML como un Script. Actualmente el lenguaje de programación es desarrollado por la empresa Sun (Sánchez, 2013).

1.1.14.1. Características

Las principales características de este lenguaje de programación son (Gauchat, 2012):

- Simple.
- Manejo de objetos.
- Dinámico.
- Funcional.
- Multiplataforma

Se puede considerar simple debido a que no se necesita amplios conocimientos en programación, maneja objetos dentro de una página web con la finalidad de definir eventos, responde a eventos en tiempo real por eso es dinámico, es funcional debido al uso de funciones predeterminadas en el lenguaje al igual de incorporar funciones creadas y multiplataforma, al ser embebido en una página HTML, puede ser ejecutado en cualquier sistema operativo.

Continua →

1.1.14.2. *Sintaxis*

Tabla 13.
Sintaxis de programación.

Características	Descripción
No cuenta espacios ni líneas nuevas	El intérprete ignora cualquier espacio en blanco sobrante.
Key Sensitive	Distinción entre mayúsculas y minúsculas.
No se define el tipo de variable	Las variables pueden almacenar diferentes tipos de datos durante la ejecución.
Embebido	Mediante el uso de etiquetas debe ser delimitado dentro del cuerpo del código HTML.
Tipo de datos soportados	<ul style="list-style-type: none"> • Números: enteros. • Cadena de caracteres: texto entre comillas. • Lógicas: valores verdaderos o falsos. • Nulos: no se da valor a la variable.

Comprende un conjunto de reglas que deben seguirse para escribir el código fuente, detallados en la tabla 13 (Eguíluz, 2008):

Existen ciertos términos que deben ser tomados en cuenta como son (Eguíluz, 2008):

- Palabras reservadas: se utiliza para construir sentencias, y no pueden ser utilizadas libremente. Son en Inglés.
- Sentencias: son las instrucciones que forman un Script.
- Script: Código creado en el lenguaje de programación.

Resumen del Software a utilizar

Se revisaron las herramientas para la realización de este aplicativo, las cuales se resumen en la tabla 14:

Tabla 14.
Detalle de Software.

Software	Productor	Versión
PHP	PHP Group	5.6
HTML	Word Wide Web Consortium	5
Java Script	Mozilla Foundation	ECMAScript 2016
Php-nlptools	Sam Hocevar	2
MySQL	Oracle Corporation	5.7.10
XAMPP	Apache	5.6.23

CAPÍTULO 3

Descripción general del proyecto

La aplicación se monta dentro de un servidor web, esta contiene scripts desarrollados en PHP que ejecutan los algoritmos para conexión con base de datos, PLN y la interacción con el usuario, todo esto se encuentra dentro de una red local (LAN), permitiendo al usuario la realización de consultas desde cualquier tipo de dispositivo vía web a la dirección asignada al servidor esquema que se muestra en la figura 7.

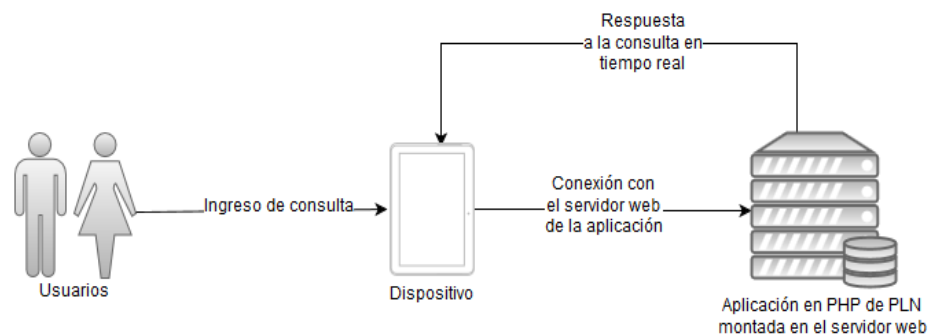


Figura 7. Esquema General del aplicativo.

Dentro de la interfaz de la página web el usuario ingresa la consulta en un cuadro texto y envía la información para que el servidor busque una respuesta al requerimiento solicitado por el usuario de existir la información dentro de su base de datos; caso contrario solicita al usuario otra consulta para ser gestionada.

Las respuestas generadas por el servidor se presentan como una impresión de texto independiente por cada sugerencia realizada, permitiendo así la realización de una sola consulta por cada envío de solicitud. El diagrama de casos se detalla a continuación la figura 8.

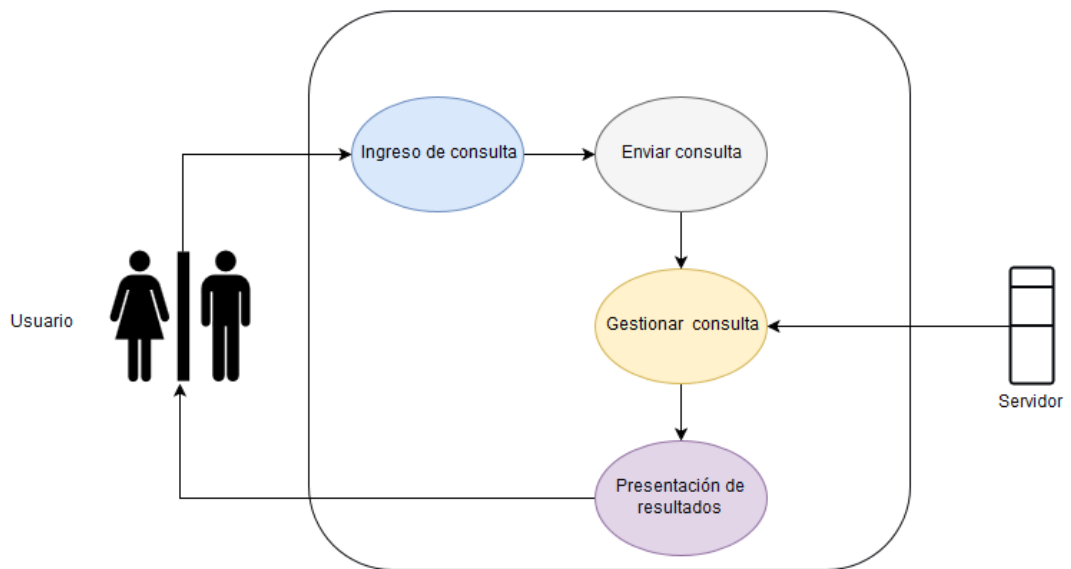


Figura 8. Diagrama de casos de la aplicación web.

Diseño del proyecto

1.1.15. Descripción del algoritmo general

Para la implementación del proyecto se realizan cinco procesos, los cuales se detallan en la figura 9.

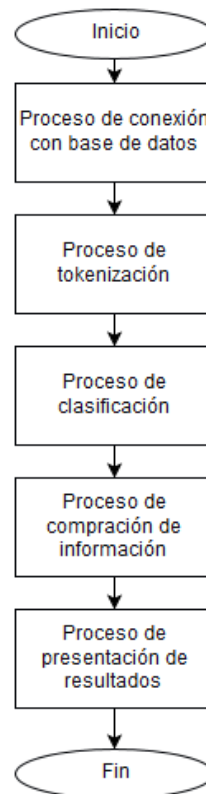


Figura 9. Descripción del algoritmo general.

El proceso inicializa al servidor web para comenzar su funcionamiento, escucha los puertos web Http, 80 y 8080; para realizar pruebas se ingresa a las direcciones locales (Http://localhost o http://127.0.0.1). Teniendo un núcleo que se comunica con cada módulo, permitiendo así el uso de APIs dentro de su estructura y permitir configurar así un servidor a la medida de los requerimientos del programador.

A continuación, se realiza la conexión a la base de datos, seguido de esto se tokeniza el texto ingresado por el usuario para ser clasificado en un tercer proceso, en una cuarta etapa se realiza la comparación de la información con la base de datos y para finalizar se presentan los resultados en el servidor web.

1.1.15.1. *Proceso de conexión con base de datos*

Dentro de la estructura de la base de datos se pretende almacenar la mayor cantidad de información respectiva a las consultas que el usuario puede presentar, es por eso que se diseñó una estructura no relación que contiene distintas tablas con la

información referente a docentes, procesos de graduación, lugares de interés del Departamento de Eléctrica y Electrónica, eventos correspondientes a la carrera y el horario de cada una de las materias dictadas por el Departamento. Esta información se detalla en la tabla 15 y en las figuras 10,11, 12,13 y 14 respectivamente.

Tabla 15.
Tablas de la base de datos.

Nombre de la tabla	Descripción
Profesor	Se lista por nombre y apellido cada uno de los docentes pertenecientes al DEEE, incluyendo el área a la que pertenece, ID, mail y horario de trabajo del profesor.
Ubicación	Se describe lugares de interés para los estudiantes del DEEE con los respectivos números telefónicos.
Graduación	Se almacena los pasos requeridos por la institución para realizar el proceso de graduación.
Eventos	Se almacena la información sobre los eventos más importantes para la institución.
Materias	Guarda la información referente a todas las asignaturas que se dictan por parte de Departamento de Eléctrica y Electrónica.

profesor	
idprofesor	INT(11)
nombre	VARCHAR(45)
apellido	VARCHAR(45)
horario	VARCHAR(45)
area	VARCHAR(45)
iD	VARCHAR(45)
mail	VARCHAR(100)

Figura 10. Descripción de la tabla profesor.

ubicacion	
idubicacion	INT(11)
ulugar	VARCHAR(200)
udireccion	VARCHAR(200)
uotros	VARCHAR(200)

Figura 11. Descripción de la tabla ubicación.

graduacion	
idgraduacion	INT(11)
gdescripcion	VARCHAR(1000)

Figura 12. Descripción de la tabla graduación.

eventos	
ideventos	INT(11)
actividad	VARCHAR(140)
fecha	VARCHAR(500)

Figura 13. Descripción de la tabla eventos.

The screenshot shows a window titled 'materias' with a dropdown arrow. Below the title, there is a list of columns with their data types and constraints:

- idmaterias INT(11) (Primary Key, indicated by a yellow lightning bolt icon)
- materianombre VARCHAR(65) (Foreign Key, indicated by a diamond icon)
- materiaprofesor VARCHAR(45) (Foreign Key, indicated by a diamond icon)
- materiahorario VARCHAR(1500) (Foreign Key, indicated by a diamond icon)

At the bottom of the window, there is a section labeled 'Indexes' with a right-pointing arrow.

Figura 14. Descripción de la tabla materias.

El proceso de conexión con la base de datos inicia con la declaración de cuatro variables, las cuales son:

- Servername: Dirección ip de la base de datos a conectar, en este caso se ocupan la dirección local host (127.0.0.1).
- Username: El usuario registrado con el cuál se permite el acceso a la base de datos.
- Password: Contraseña asignada al usuario registrado.
- Dbname: Nombre de la base de datos a la cual se realiza la conexión.

A continuación, mediante la integrabilidad de MySQLy PHP, con el uso de la función “conn” declarada dentro de la clase “mysqli” disponible dentro de la librería de PHP; se comprueba que las variables ingresadas sean correctas, de serlo se procede al siguiente proceso con una conexión exitosa; caso contrario se presenta un error de conexión y todos los procesos siguientes son cancelados. Esta función se detalla en la figura 15 y el esquema de conexión en la figura 16.

```
$conn = new mysqli($servername, $username, $password, $dbname);
// Check connection
if ($conn->connect_error) {
    die("Connection failed: " . $conn->connect_error);
}
```

Figura 15. Código de conexión a base de datos.

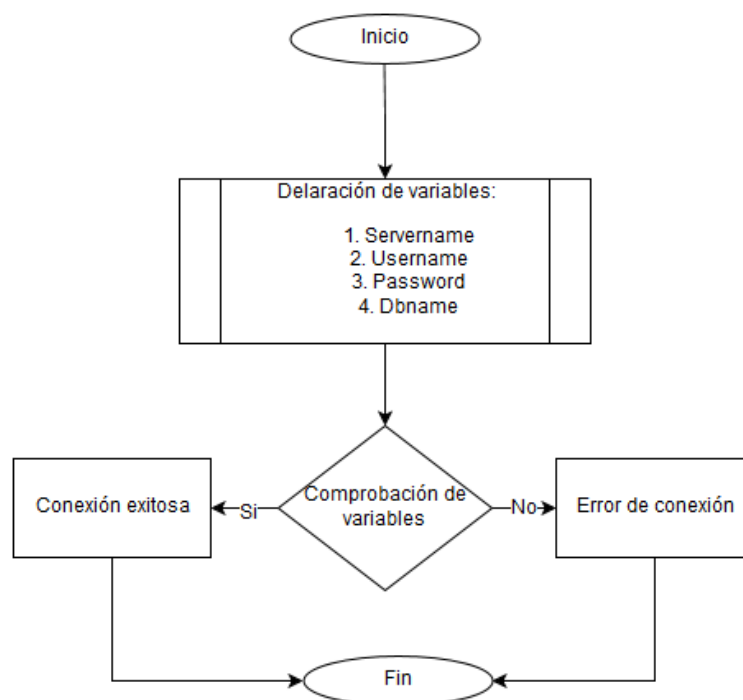


Figura 16. Proceso de conexión con base de datos.

1.1.15.2. Proceso de tokenización

Lo primero que se realiza en este proceso es la importación de la librería:

- NlpTools\Tokenizers\WhitespaceAndPunctuationTokenizer

La cual contiene la información referente a todos los procesos de tokenización (*Tokenizers*); la siguiente subclase indicada que cada signo de puntuación es considerado un token independiente, no incluye espacios en blanco como un token (*WhitespaceAndPunctuationTokenizer*).

Se declara un objeto de la clase “*WhitespaceAndPunctuationTokenizer*”, se ingresa el texto a tokenizar usando la función “tokenize” y el resultado es almacenado en una variable auxiliar para los siguientes procesos. La función se muestra en la figura 17 y el proceso en la figura 18.

```

$tok = new WhitespaceAndPunctuationTokenizer();
$a=$tok->tokenize($_POST['apellido']);
  
```

Figura 17. Función de tokenización.

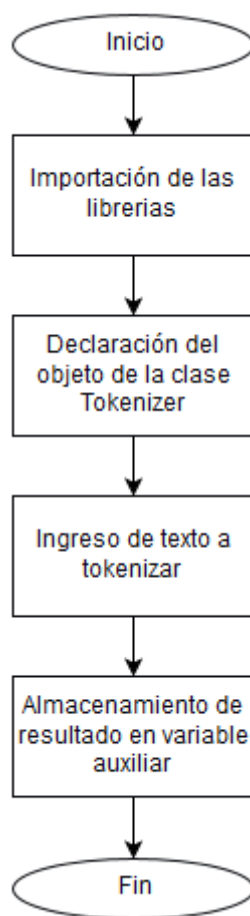


Figura 18. Proceso de tokenización.

1.1.15.3. *Proceso de clasificación*

El proceso empieza con la importación de la librería:

- NlpTools\Similarity\JaccardIndex

La clase “Similarity” contiene subclases referentes a la comparación y búsquedas de similitudes de vectores tokenizados, la subclase “JaccardIndex” calcula el índice de Jaccard, ya explicado en el capítulo 2 sección 2.3.2.1.

Mediante la declaración de un objeto de clase JaccardIndex, se comparó la similaridad del texto obtenido el proceso de tokenización con un conjunto de parámetros de clasificación, para el caso de la tabla Profesores; se filtran los apellidos de los docentes del Departamento de Eléctrica y Electrónica.

Posteriormente se procedió a realizar el cálculo del índice ya mencionado; si este es mayor que cero el resultado se lo almacena en una variable auxiliar para los procesos siguientes, caso contrario se cambia el conjunto de parámetros de clasificación. Esta función se muestra en la figura 19 y el proceso en la figura 20.

```
$tok2= new JaccardIndex();
$dAlulema=$tok2->similarity($a,$auxAlulema);
```

Figura 19. Función de cálculo de índice de Jaccard.

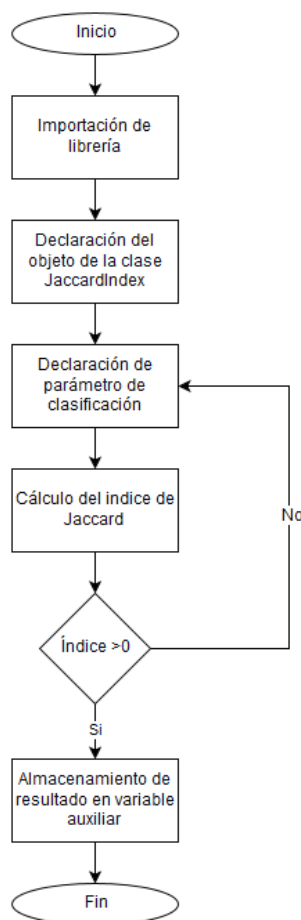


Figura 20. Proceso de clasificación.

1.1.15.4. *Proceso de comparación de información*

Se declaran un conjunto de parámetros complementarios de comparación como son detallados en la tabla 16:

Tabla 16.
Parámetros de comparación de información.

Parámetro	Descripción
Horario	Contiene la información del horario del docente.
Área	Contiene el área de trabajo del docente.
ID	Contiene el Identificador de cada profesor.
Mail	Almacena el correo electrónico de cada docente.
Extra	Detalla toda la información referente al docente.

Se declara un nuevo índice de Jaccard y se lo calcula individualmente con cada uno de los parámetros mencionados, si este valor es mayor que cero se procede a la extracción de información de la base de datos; caso contrario se cambia el parámetro de consulta. Este proceso es detallado en la figura 21.

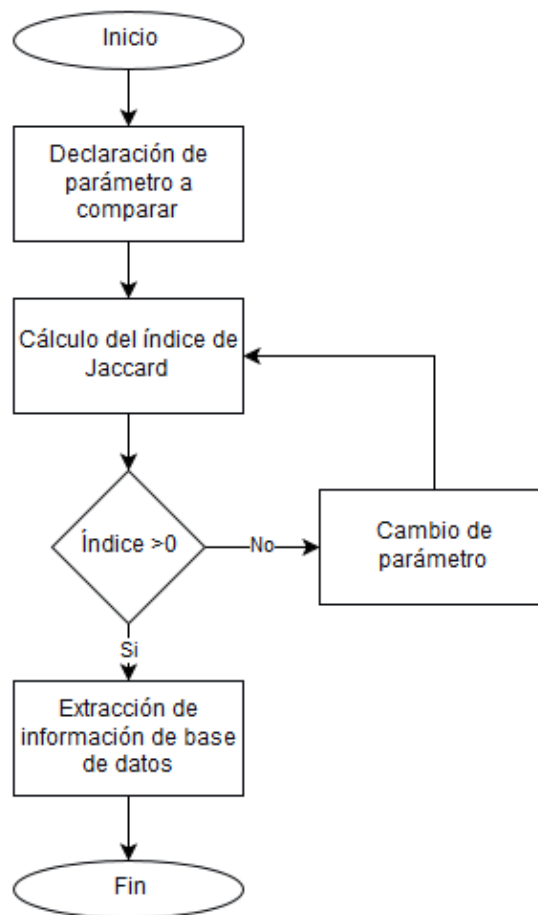


Figura 21. Proceso de comparación de información.

1.1.15.5. Proceso de presentación de resultados

Se da tratamiento a todos los resultados de procesos anteriores, dando un formato acorde al chatbot conversacional para ser presentado en pantalla en el menor tiempo posible, brindando así una sensación de respuestas en tiempo real. Este proceso se detalla en la figura 22.

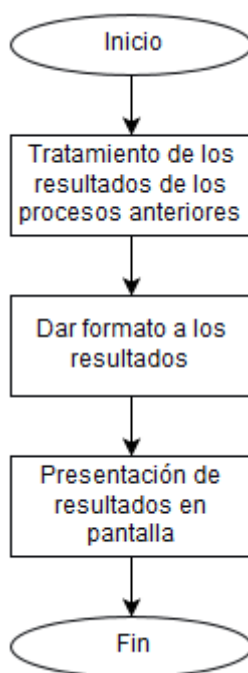


Figura 22. Proceso de impresión.

Descripción de requisitos

Los requisitos de la aplicación se detallan en la tabla 17.

Tabla 17.
Descripción de requisitos.

Requisito	Descripción
Procesamiento del Lenguaje Natural	Esto se realiza mediante el uso de la herramienta Npltools escrita para PHP.
Configuración de la base de datos	Modelamiento y estructuración de una base de datos que contiene la información del Departamento de Eléctrica y Electrónica, configurado en MySQL.
Montaje del servidor web	Búsqueda de servidores online que permitan la expansión de la aplicación web de un servidor local al internet

Diseño de interfaz web

Dentro de la estructura de la página web, se la puede dividir en dos bloques; uno que contiene la cabecera, donde se declara las importaciones de las librerías, se realiza la conexión a la base de datos y la implementación de los formularios; el segundo contiene el cuerpo de la aplicación web, es decir todos los procesos declarados en las secciones anteriores. Este esquema se detalla en la figura 23.



Figura 23. Esquema de la página web.

Consideraciones de diseño

1.1.16. Diseño del Filtro para tabla profesores

La tabla profesores posee diversos campos o atributos con los cuales un usuario puede realizar la búsqueda o el filtrado correspondiente en función de la información que desea consultar, para esto se realiza cuatro pruebas de diseño que se resumen en la tabla 18; estas pruebas consisten en verificar y eliminar errores que generen confusión o ambigüedad al sistema diseñado; estas son: la primera un filtro diseñado en base al nombre del profesor, la segunda en función del apellido, la tercera en función del mail y la última por su ID.

Tabla 18.
Diseño de filtros para la tabla profesores.

Prueba/Filtro	Resultado
Nombre	Se genera alto número de ambigüedad debido a que para muchos docentes el nombre común.
Apellido	Presenta un índice de ambigüedad menor que el filtro diseñado respecto a nombres, debido a los apellidos son poco repetitivos.
Mail	Si bien el mail es la unión del nombre y apellido del docente, puede generarse más de un mail similar, presentando diferencias con un número complementario para su distinción
ID	No es muy popular ni conocido este parámetro para consultas por los estudiantes.

De los resultados obtenidos, no es conveniente el uso del mail y del ID debido a que no son tan comunes para su uso por parte de los estudiantes y profesores, por estos filtros son descartados para el acceso principal a los datos del docente, pero si son considerados como palabras claves para las búsquedas internas de información.

Con respecto a los filtros de nombre y apellido, el primero debido en este caso al gran número de repeticiones existentes entre los docentes se debería crear un filtro extra para acceder a la información completa; mientras el bajo número de ambigüedad que se presenta en los apellidos hace que sea más sencillo el acceso a los datos y en caso de ambigüedad al ser reducidos pueden ser filtrados mediante el uso de una comprobación extra.

1.1.17. Uso del índice de Jaccard para el proceso de filtrado


Con el índice de Jaccard se basa el proceso de filtrado para la búsqueda en cada categoría, pero esto es diferente en el acceso a cada tabla en la base de datos; es por eso que se determinaron pruebas para ver el nivel de características que pueden ser consultadas, esto se detalla en la tabla 19.

Para el caso de los profesores se crea un conjunto de datos, los cuales son los apellidos de cada uno los profesores como un grupo general, basándose en el modelo booleano de recuperación de información del PLN, en el cual se busca la presencia o usencia del parámetro a filtrar. De existir el índice de Jaccard cambia o incrementa su valor en función de toda la sentencia tokenizada, accediendo a otro conjunto que contiene las palabras de filtraje para acceder a los datos respetivos de cada docente, esto se realiza semejante al modelo probabilístico de recuperación de información ya que se calcula el índice de Jaccard de cada una de estas palabras en la sentencia tokenizada.

Para la tabla ubicación se pretende filtrar el nombre del lugar buscando la presencia del mismo, dentro de una agrupación con información referente a la posición geográfica de cada sitio a ser consultado; mismo proceso que se realiza para la tabla del proceso de graduación y eventos, como se detallan en las figuras 24, 25, 26, 27 y 28 respectivamente.

Tabla 19.
Uso del índice de Jaccard en el P.F.

Tabla de base de datos	Descripción de proceso
Profesores	Se requiere de cálculo de un índice de Jaccard general para el acceso a la información de docente y otros anidados al mismo para permitir la consulta de la información requerida.

Continua 

Ubicación	El planteamiento realizado es el cálculo de un solo índice de Jaccard que filtra el nombre del lugar consultado.
Graduación	Debido a la forma del proceso de graduación y la relación relacionada a esto, solo se realiza el cálculo de un índice de Jaccard para filtrar en general todo lo relacionado al proceso.
Eventos	Se realiza un cálculo del índice de Jaccard para filtrar la información referente a esta categoría.
Materias	Para esta tabla se realiza el cálculo de un índice de Jaccard para acceder al nombre de la asignatura y desplegar toda la información.

Debido a que la tabla profesores que posee más campos de consultas es necesario la realización de un Jaccard principal para seleccionar al docente y otros anidados a este para cada uno de los campos de las sugerencias, este proceso se detalla en la figura 24. Para las otras tablas no es necesario anidar cálculos de índice de Jaccard; esto por el bajo número de componentes en sus tablas.

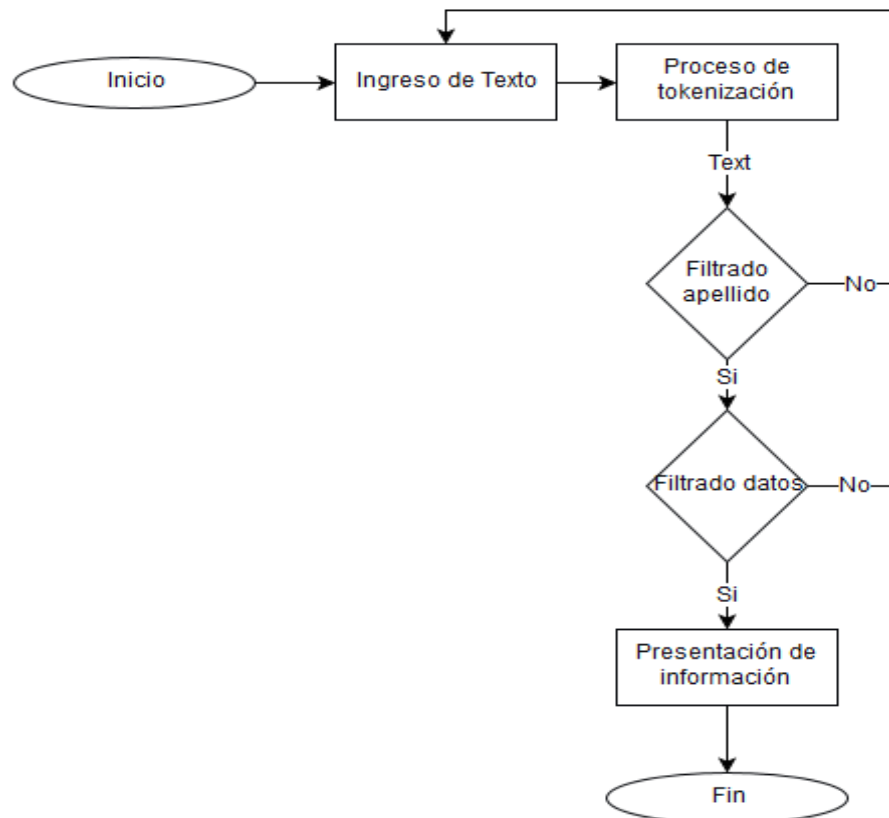


Figura 24. Anidación de Jaccard para tabla profesores.

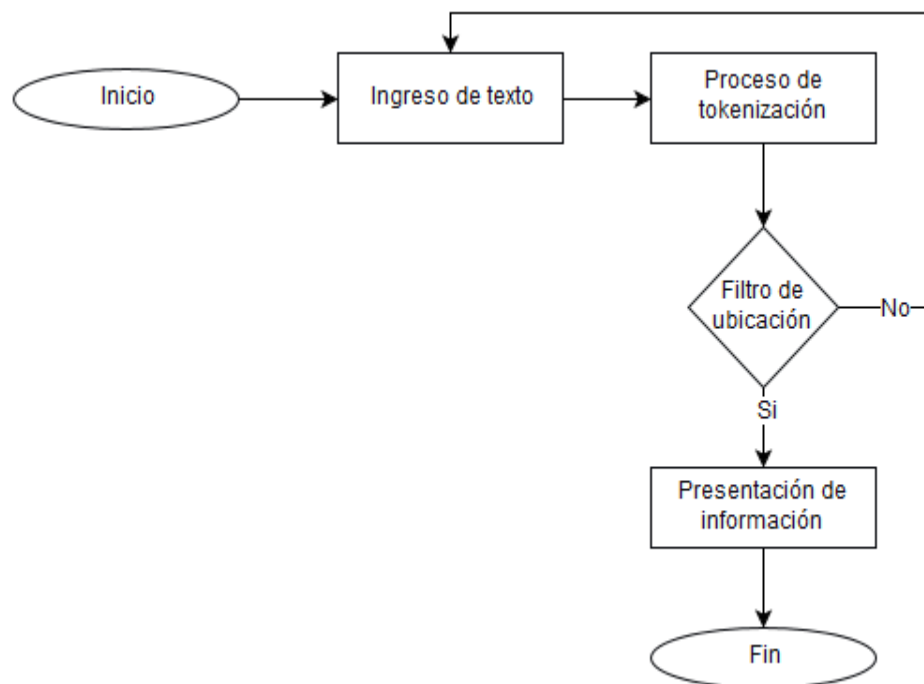


Figura 25 . Anidación de Jaccard para tabla ubicación.

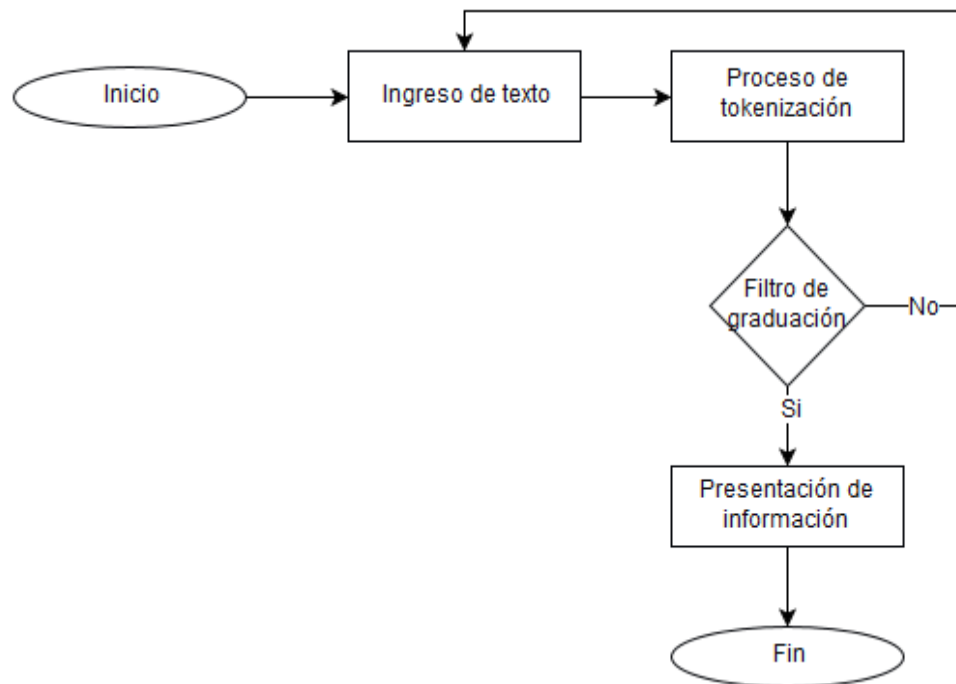


Figura 26. Anidación de Jaccard para tabla graduación.

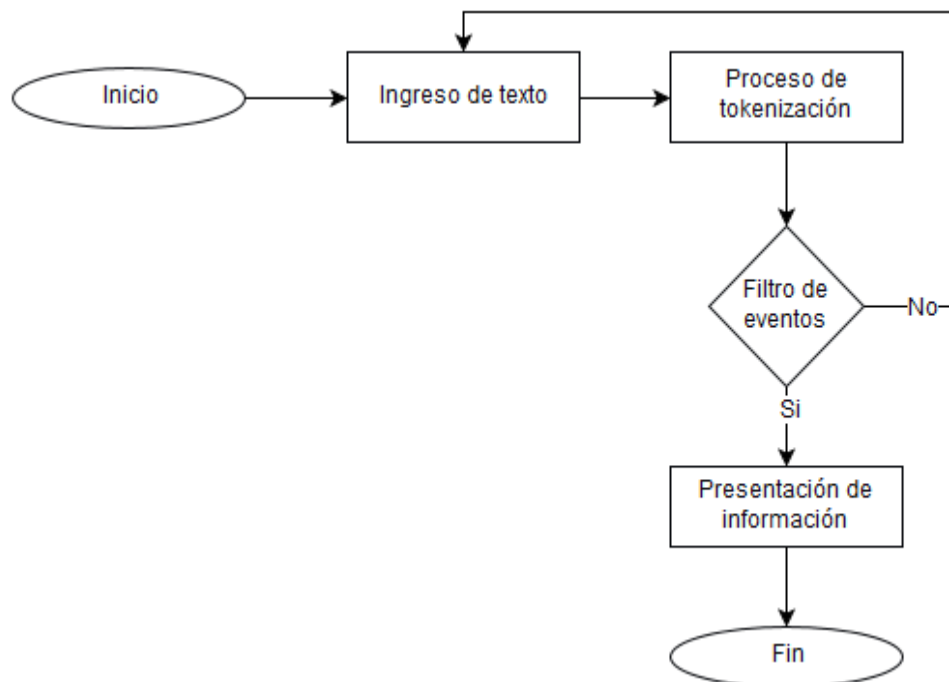


Figura 27. Anidación de Jaccard para tabla eventos.

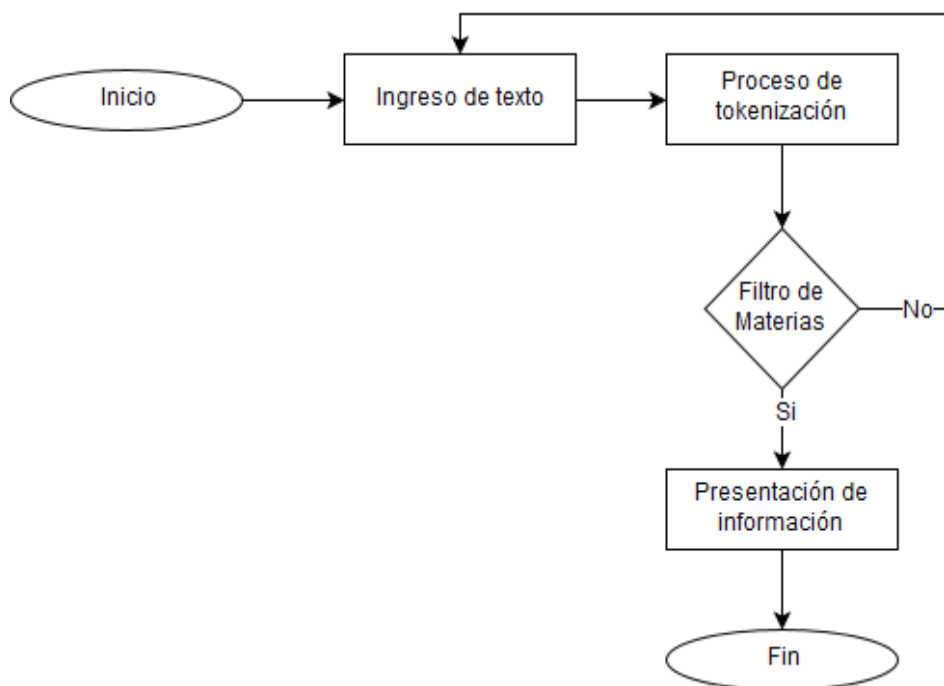


Figura 28. Anidación de Jaccard para tabla materias.

1.1.18. Métodos de ingreso de texto

Para estas consideraciones de diseño se plantearon dos escenarios: el primero es la lectura online del texto ingresado, esto consiste en adquirir los datos mientras el usuario está escribiendo y el segundo escenario consiste en que el usuario ingrese la consulta y posterior se presiona el botón enviar, es detallado en la tabla 20.

Tabla 20.
Comparación de métodos de ingreso de texto.

Escenario	Resultado
Lectura online	Se presenta una dificultad al momento de capturar los datos debido a que no se puede predecir la intención de consulta del usuario.
Botón enviar	Debido al uso de botón enviar ya se posee toda la sentencia con toda la información que el usuario desea consultar, facilitando el proceso de filtrado.

Por medio de la facilidad y el uso de los formularios en PHP, se prefiere el uso del método de ingreso de texto y posterior presión del botón enviar.

1.1.19. Manejo de caracteres especiales

La versión actual del proyecto en PHP tiene un soporte para caracteres especiales limitado, utilizando la codificación UTF-8 permitiendo así el uso de tildes y “ñ” dentro de la programación y el ingreso del texto, aunque a veces presenta errores que pueden ser solventados mediante la conversión de texto o supresión de caracteres especiales. Por otro lado MySQL posee palabras reservadas y un limitado manejo de caracteres especiales; es por eso que al unir estos dos lenguajes se requiere la unificación del manejo de los caracteres especiales por lo que se opta por mantener el esquema de MySQL no permitiendo el uso de la ñ, tildes y la comilla simple, aparte los mails y ciertos datos de los docentes suprimen caracteres especiales, contribuyendo así a la decisión tomada, se describe en la tabla 21.

Tabla 21.
Conflicto de caracteres especiales en diferentes lenguajes.

Carácter especial	Conflicto
Ñ	Caracter que afecta al lenguaje PHP sin utilizar codificación UTF-8
‘(comilla simple)	Instrucción reservada en MySQL
Tildes	Caracteres que afectan al lenguaje PHP sin utilizar codificación UTF-8

CAPÍTULO 4

Introducción

En este capítulo, se desarrollan distintos escenarios de prueba con los cuales se pretende evaluar el desempeño íntegro del aplicativo con el objetivo de analizar el comportamiento del sistema y plantear futuras mejoras.

Escenario de prueba

1.1.20. Escenario 1: Determinación de la eficacia del algoritmo

En el capítulo 3 sección 3.5 se diseñó un algoritmo capaz de clasificar las consultas en función de un conjunto de palabras que mantienen una relación con cada una de las tablas dentro de la base de datos.

Para este escenario, la determinación de la eficiencia viene dada con la realización de consultas en base a las preguntas más comunes presentadas durante el proceso de diseño dentro de cada categoría; para así cuantificar dentro de un esquema si es correcto o no el contexto de la respuesta que presenta el sistema.

Cada pregunta será evaluada bajo un criterio: donde “3” implica la respuesta esperada, “2” que contiene algo de la información solicitada o repetitiva y “1” siendo respuesta incorrecta.

1.1.20.1. *Tabla Profesores*

Para la búsqueda de docentes se determinaron tres pruebas que engloban las consultas más comunes que pueden suscitarse, estas se describen en la tabla 22.

Tabla 22.
Descripción de las pruebas para la tabla Profesores.

Número de prueba	Tema
Prueba 1	Consulta de datos específicos dentro de la tabla en la base de datos.

Continua

Prueba 2	Consulta de todos los datos de la tabla.
Prueba 3	Consulta de ambigüedad.

Para la Prueba 1 se describen las siguientes preguntas:

- ¿Cuál es el área de trabajo del ingeniero Alulema?
- ¿Cuál es el área y el ID del inge Bernal?
- ¿Cuál es el ID, mail y el área del trabajo de la inge Nancy Paredes?

Para la Prueba 2 se describen las siguientes preguntas:

- ¿Quién es el Ingeniero Diego Villamarin?
- ¿Quién es la Ingeniera Zapata?
- ¿Me puedes dar información del Ing Alvarez?

Para la Prueba 3 se describen las siguientes preguntas:

- ¿Quién es el ingeniero De la Cruz?
- ¿Quién es el ingeniero Acosta?
- ¿Me podrías ayudar con la información del Ing Aguilar?

Los resultados para las pruebas 1, 2 y 3 se describen en las tablas 23, 24 y 25 respectivamente.

Para la prueba 1, cuando se realizó las preguntas que contienen más de dos parámetros de consulta el resultado es redundante o repetitiva por lo cual se calificó como regular, para los casos donde solo se realizó una consulta específica el resultado fue lo esperado.

Tabla 23.
Resultados de Prueba 1.

Número de pregunta	Resultado
Pregunta 1	3
Pregunta 2	2
Pregunta 3	2

Para la prueba 2, los resultados fueron los esperados ya que el sistema proporcionó la información completa de los docentes consultados.

Tabla 24.
Resultados de Prueba 2.

Número de pregunta	Resultado
Pregunta 1	3
Pregunta 2	3
Pregunta 3	3

Para la prueba 3, debido a que en el Departamento de Eléctrica y Electrónica existen apellidos de docentes repetidos se generó la impresión informativa de los mismos.

Tabla 25.
Resultados de Prueba 3.

Número de pregunta	Resultado
Pregunta 1	2
Pregunta 2	2
Pregunta 2	2

Los resultados generales para la tabla Profesores se muestran en las figuras 29 y 30, la eficiencia del algoritmo de clasificación se describe en la tabla 26.

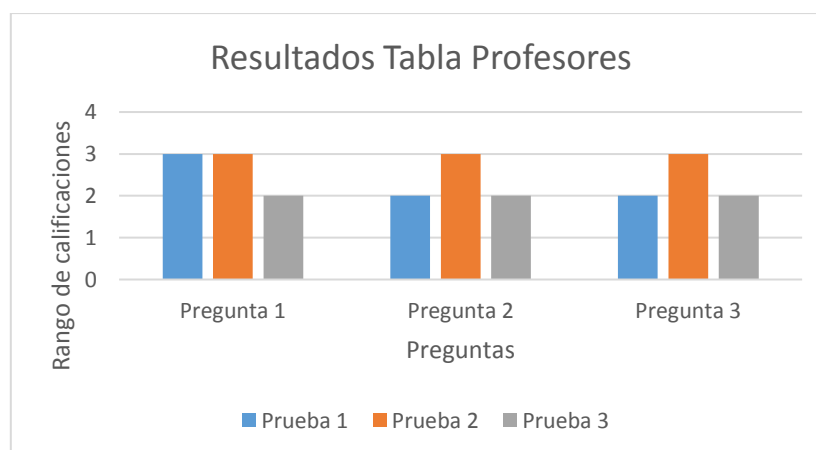


Figura 29. Resultados tabla Profesores.

De los datos se pudo observar que no existe un promedio de preguntas mal contestadas por el sistema y contiene una mayoría de respuestas regulares, lo que implica que este porcentaje no involucra errores del aplicativo por lo cual se pueden agrupar para el cálculo de la eficiencia como un porcentaje bueno; es decir que las nueve preguntas realizadas para este tipo de consultas el algoritmo presenta una eficiencia superior al 80% y el error porcentual que se presenta es de ambigüedad que fueron corregidos después de la realización de esta prueba.

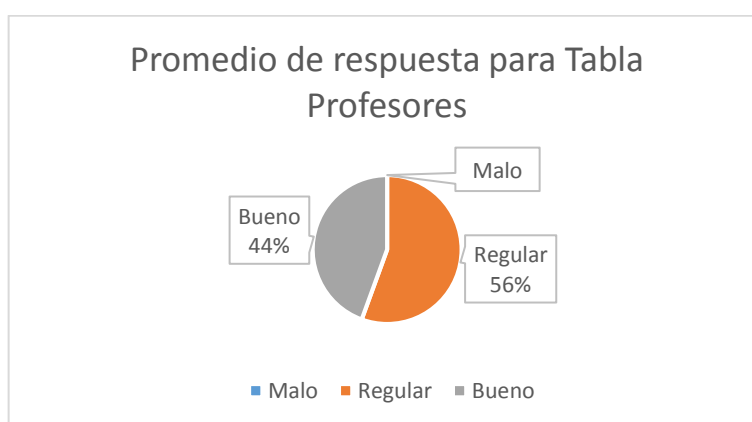


Figura 30. Promedio de respuesta para tabla Profesores.

Tabla 26.
Eficiencia del algoritmo para la tabla Profesores.

Eficiencia (%)	83,333335
Error (%)	16,666665

1.1.20.2. Tabla Materias

Para la búsqueda de materias dictadas por el Departamento de Eléctrica y Electrónica se determinaron dos pruebas que engloban las consultas generales que pueden suscitarse, estas se describen en la tabla 27.

Tabla 27.
Descripción de las pruebas para la tabla Materias.

Número de prueba	Tema
Prueba 1	Consulta de todos los datos de tabla.
Prueba 2	Consulta de ambigüedad.

Para la Prueba 1 se describen las siguientes preguntas:

- ¿Cuál es el horario de la materia de electrofluidos?
- ¿Cuál es el horario de la materia de NGN?

Para la Prueba 2 se describen las siguientes preguntas:

- ¿Cuál es el horario de la materia de electrónica?
- ¿Cuál es el horario de la materia de electromagnetismo?

Los resultados para las pruebas 1 y 2 se describen en las tablas 28 y 29 respectivamente.

Para la prueba 1, los resultados fueron los esperados ya que el sistema proporcionó la información completa sobre la materia consultada. Como se detalla en la tabla 28.

Tabla 28.
Resultados de Prueba 1.

Número de pregunta	Resultado
Pregunta 1	3
Pregunta 2	3

Para la prueba 2, existió ambigüedad al consultar materias ya que su estudio es consecutivo. El resultado se detalla en la tabla 29.

Tabla 29.
Resultados de Prueba 2.

Número de pregunta	Resultado
Pregunta 1	2
Pregunta 2	2

Los resultados generales para la tabla Materias se muestran en las figuras 31 y 32, la eficiencia del algoritmo de clasificación se describe en la tabla 30.

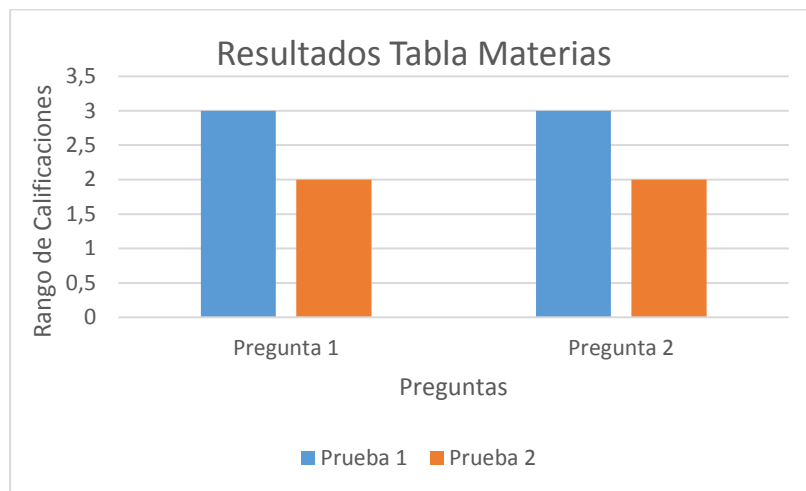


Figura 31. Resultados Tabla Materias.

De los datos obtenidos se puede visualizar que no existe un promedio de preguntas mal contestadas por el sistema y existe igual cantidad de respuestas regulares como buenas, lo que implica que este porcentaje no involucra errores del aplicativo por lo cual se pueden agrupar para el cálculo de la eficiencia como un porcentaje bueno; es decir que las cuatro preguntas realizadas para este tipo de consultas el algoritmo presenta una eficiencia superior al 80% y el error porcentual que se presenta es de ambigüedad.

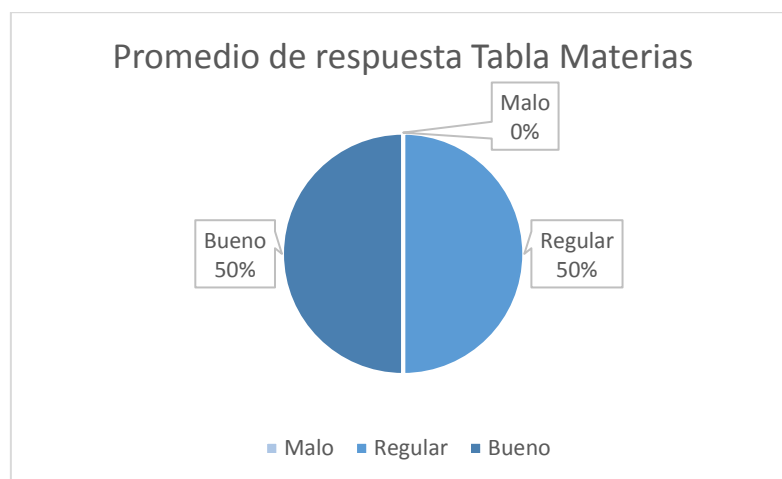


Figura 32. Promedio de respuesta Tabla Materias.

Tabla 30.
Eficiencia del algoritmo para la tabla Materias.

Eficiencia (%)	81,4814815
Error (%)	18,5185185

1.1.20.3. *Tabla Ubicación*

Para la búsqueda de direcciones dentro de la Universidad de las Fuerzas Armadas ESPE se determinó una prueba que engloba las consultas frecuentes que pueden suscitarse, estas se describen en la tabla 31.

Tabla 31.
Descripción de las pruebas para la tabla Ubicación.

Número de prueba	Tema
Prueba 1	Consulta de todos los datos de tabla.

Para la Prueba 1 se describen las siguientes preguntas:

- ¿Dónde queda la ESPE?
- ¿Dónde está ubicada la UAR?
- ¿Dónde está el laboratorio de Telecomunicaciones?

Para la prueba 1, los resultados fueron los esperados ya que el sistema proporcionó la información completa de la ubicación de los lugares consultados. Como se detalla en la tabla 32.

Tabla 32.
Resultados de Prueba 1.

Número de pregunta	Resultado
Pregunta 1	3
Pregunta 2	3
Pregunta 3	3

Los resultados generales para la tabla Ubicación se muestran en las figuras 33 y 34, la eficiencia del algoritmo de clasificación se describe en la tabla 33.

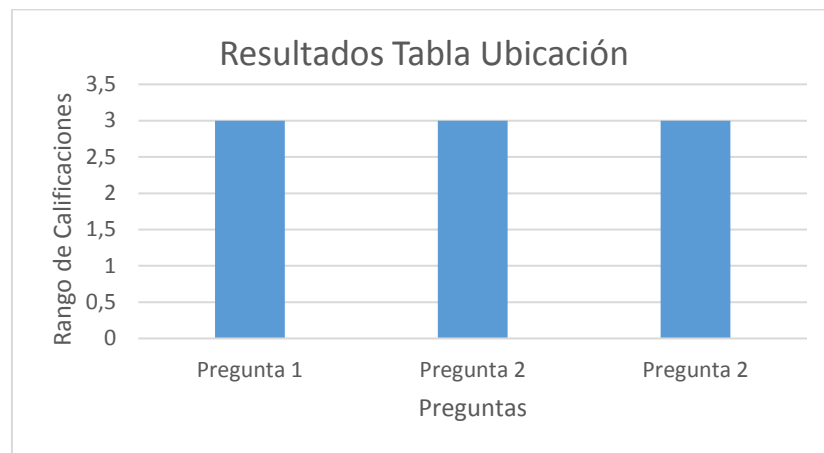


Figura 33. Resultados tabla Ubicación.

De las tres preguntas realizadas, los resultados fueron satisfactorios lo que da por resultado que la eficiencia del algoritmo para este tipo de consultas sea del 100%.

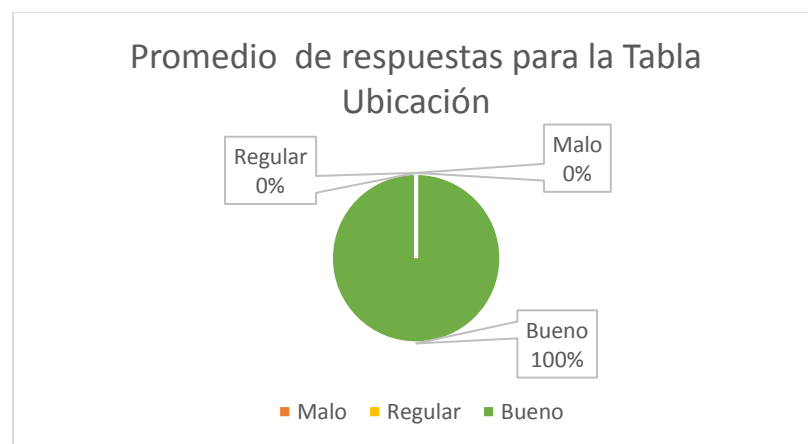


Figura 34. Promedio de respuesta para tabla Ubicación

Tabla 33. Eficiencia del algoritmo para la tabla Ubicación.

Eficiencia (%)	100
Error (%)	0

1.1.20.4. *Tabla Eventos*

Para la búsqueda de eventos de la Universidad de las Fuerzas Armadas ESPE se determinó una prueba que engloba las consultas frecuentes que pueden producirse, estas se describen en la tabla 34.

Tabla 34.
Descripción de las pruebas para la tabla Eventos.

Número de prueba	Tema
Prueba 1	Consulta de todos los datos de tabla.

Para la Prueba 1 se describen las siguientes preguntas:

- ¿Cuándo es el inicio de clases?
- ¿Cuándo hay feriado?
- ¿Cuál es último día de clases?

Para la prueba 1, los resultados fueron los deseados ya que el sistema proporcionó la información completa sobre los diferentes eventos consultados. Como se detalla en la tabla 35.

Tabla 35.
Resultados de Prueba 1.

Número de pregunta	Resultado
Pregunta 1	3
Pregunta 2	3
Pregunta 3	3

Los resultados generales para la tabla Eventos se muestran en las figuras 35 y 36, la eficiencia del algoritmo de clasificación se describe en la tabla 36.

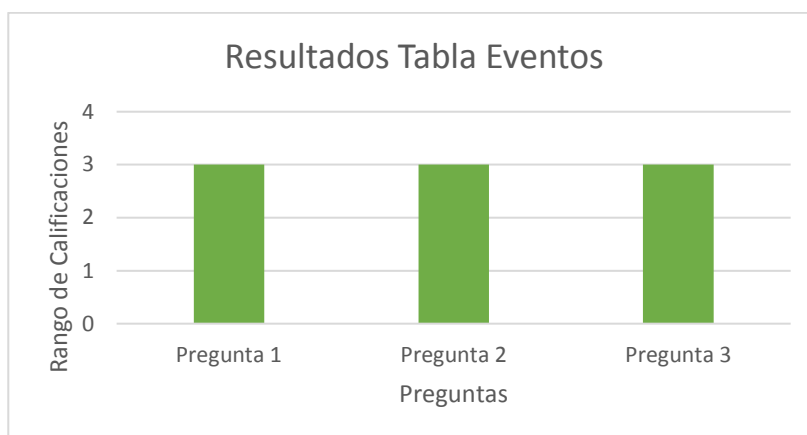


Figura 35. Resultados tabla Eventos

De las tres preguntas realizadas, los resultados fueron satisfactorios lo que da por resultado que la eficiencia del algoritmo para este tipo de consultas sea del 100%.

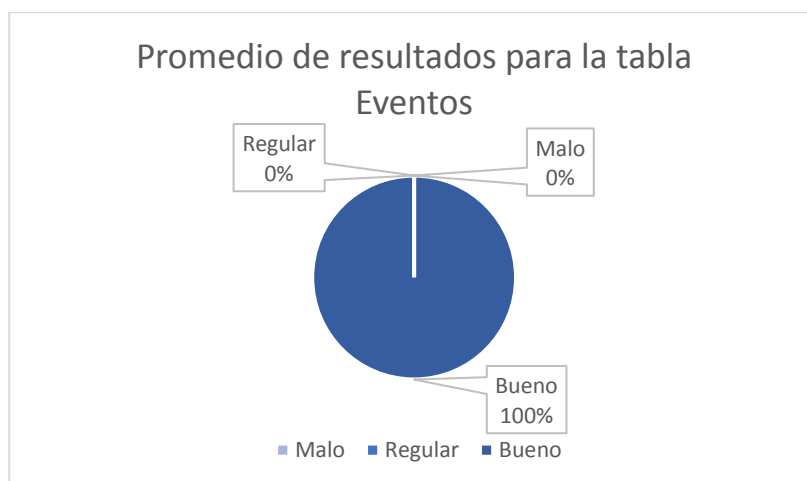


Figura 36. Promedio de respuesta para tabla Eventos.

Tabla 36.
Eficiencia del algoritmo para la tabla Eventos.

Eficiencia (%)	100
Error (%)	0

1.1.20.5. Tabla Graduación

Para la búsqueda del proceso de graduación en la Universidad de las Fuerzas Armadas - ESPE se determinó una prueba que engloba las consultas usuales que pueden producirse, estas se describen en la tabla 37.

Tabla 37.
Descripción de las pruebas para la tabla Eventos.

Número de prueba	Tema
Prueba 1	Consulta de todos los datos de tabla.

Para la Prueba 1 se describen las siguientes preguntas:

- ¿Cuáles son los requisitos para el proceso de graduación?
- ¿Cómo se evalúa el grado?
- ¿Cómo se realiza el seguimiento de graduación?

Para la prueba 1, los resultados fueron los esperados ya que el sistema proporcionó la información completa sobre los procesos consultados. Como se detalla en la tabla 38.

Tabla 38.
Resultados de Prueba 1.

Número de pregunta	Resultado
Pregunta 1	3
Pregunta 2	3
Pregunta 3	3

Los resultados generales para la tabla Graduación se muestran en las figuras 37 y 38, la eficiencia del algoritmo de clasificación se describe en la tabla 39.

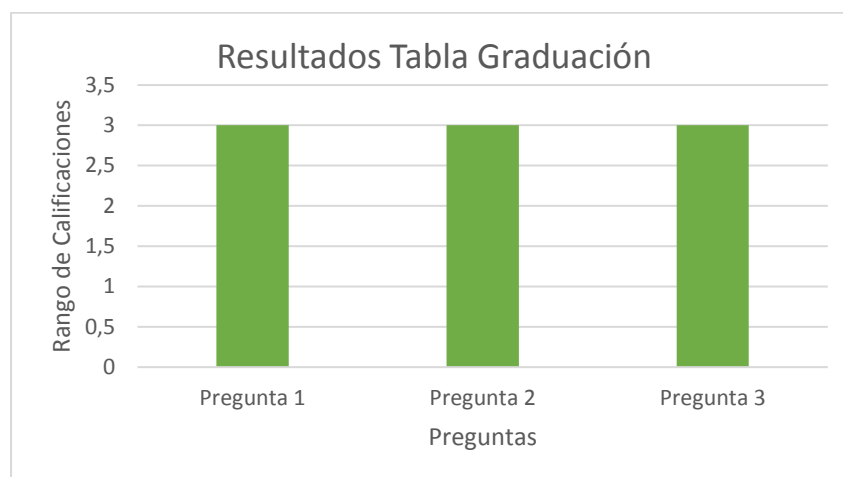


Figura 37. Resultados tabla Graduación.

De las tres preguntas realizadas, los resultados fueron satisfactorios lo que da por resultado que la eficiencia del algoritmo para este tipo de consultas sea del 100%.

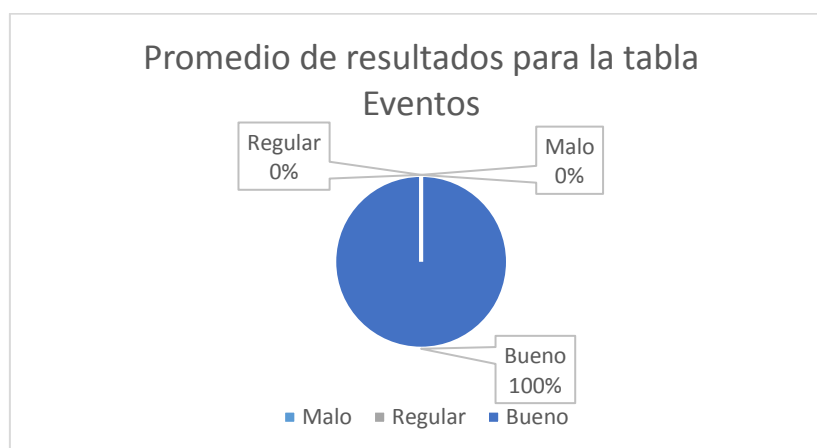


Figura 38. Promedio de respuesta para la tabla Graduación.

Tabla 39.
Eficiencia del algoritmo para la tabla Graduación.

Eficiencia (%)	100
Error (%)	0

1.1.20.6. Eficiencia Total

Para el cálculo de la eficiencia total del algoritmo, se promediaron todas las eficiencias obtenidas durante las pruebas realizadas, se muestra en la tabla 40.

Tabla 40.
Eficiencia total del algoritmo.

Eficiencia de algoritmo para tabla profesores (%)	81.48
Eficiencia de algoritmo para tabla materias (%)	83.33
Eficiencia de algoritmo para tabla ubicación (%)	100
Eficiencia de algoritmo para tabla eventos (%)	100
Eficiencia de algoritmo para tabla graduación (%)	100
Eficiencia total del algoritmo (%)	92.962
Error (%)	7.038

Debido a que el algoritmo supera el 92% de eficiencia, se lo consideró ideal para el proyecto.

1.1.21. Escenario de pruebas 2: Pruebas de servidor

1.1.21.1. Servidor local

Se diseñó un escenario en el cuál el aplicativo se lo ejecutó dentro de una red LAN privada. Las principales características de la prueba se detallan en la tabla 41 y la estructura del escenario se muestra en la figura 39.

Tabla 41.
Tipos de prueba en el servidor local.

Tipo de prueba	Resultado obtenido
Levantamiento del servidor	No se presentaron errores de conexión en el puerto 8080 donde se ejecuta la aplicación.
Conexión a base de datos	No se generó error de la conexión a la base de datos local.
Ejecución en navegador web	El servidor se ejecutó de la manera esperada.

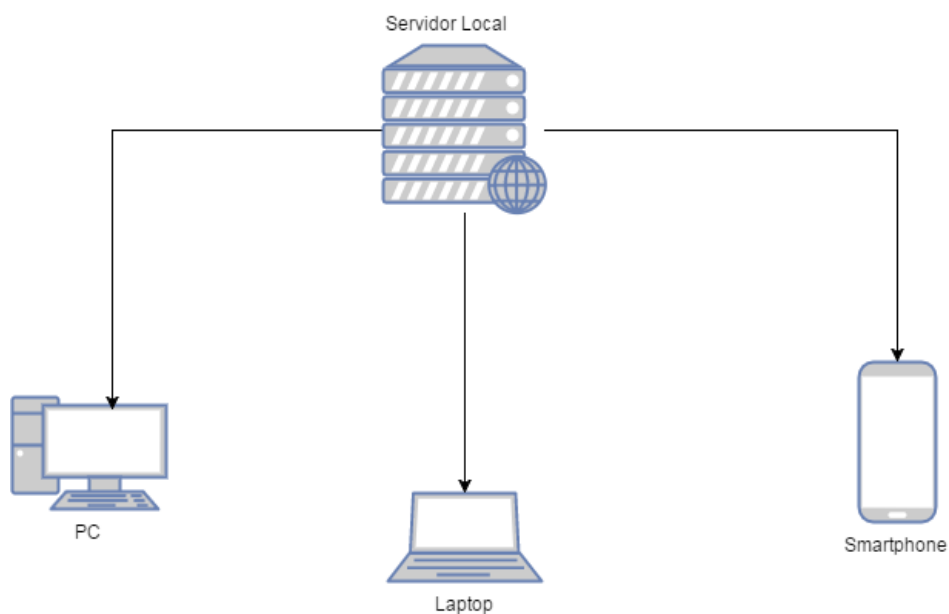


Figura 39. Estructura del escenario de pruebas del servidor local.

En la figura 40, se observa la ejecución de la aplicación dentro la red local.



Figura 40. Ejecución de la aplicación en LAN.

En la figura 41, después de realizar algunas preguntas se puede visualizar el historial a modo de chat que se almacena en un documento externo el cuál el sistema lee después de la realización de cada pregunta.



Figura 41. Historial del chat.

1.1.21.2. Servidor Externo

Para la realización de este escenario de pruebas el primer paso fue la selección de un hosting adecuado para el servidor, este proceso se describe en la tabla 42 y la estructura del escenario se muestra en la figura 42.

Tabla 42.
Hosting para el servidor externo.

Nombre del hosting	Costo	Observaciones
Wordpress.com	Gratuito	No permite la inserción de código PHP.
Wordpress.org	Gratuito	El dominio para el servidor web tiene costo.
200mb.net	Gratuito	Limitado ancho de banda, intermitencias en el despliegue del aplicativo.
Mipropia.com	Gratuito	Errores en la codificación UTF-8.

Del hosting descrito anteriormente, el que más se adapta a la necesidad del servidor y que permite que las respuestas sean en tiempo real fue: “mipropia.com”,

para la corrección del error de codificación se agregó un script el cual permite el uso de UTF-8 dentro de la aplicación.

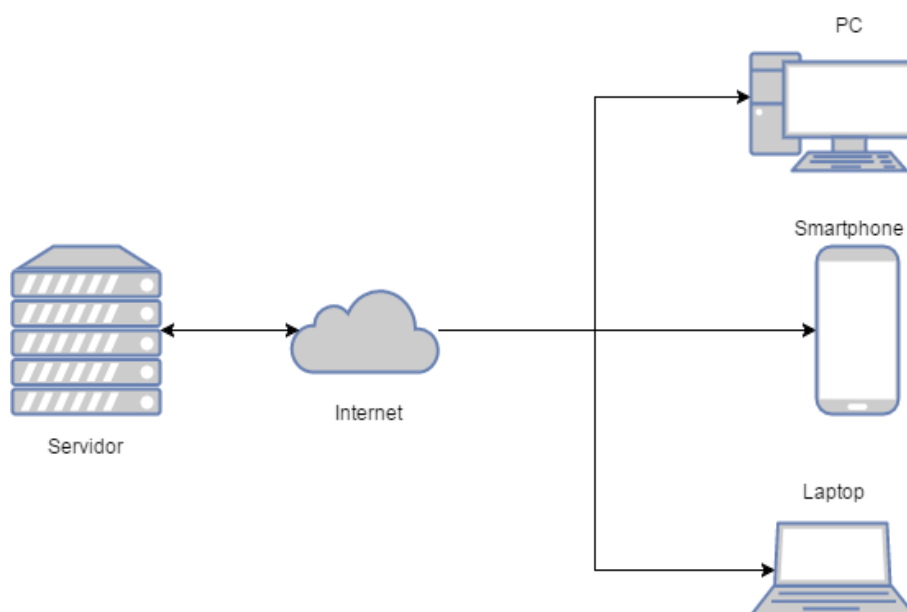


Figura 42. Estructura del escenario de pruebas del servidor externo.

En la figura 43, se observa la ejecución de la aplicación dentro del servidor externo.



Figura 43. Ejecución del aplicativo en el servidor externo.

En la figura 44, se describe la respuesta de una pregunta, el hosting seleccionado no permite conexiones ftp dentro de los scripts a ejecutarse por lo cual limita la creación de archivos adicionales evadiendo la implementación del historial del chat.



Figura 44. Respuesta de consulta en servidor externo.

1.1.22. Escenario de Prueba 3: Visualización

La realización de las pruebas de visualización se llevó a cabo en los dispositivos para el acceso a internet más comunes, es decir: un computador, un Smartphone y tablet con diferentes tamaños de pantalla, sistemas operativos y versiones de navegador, se detalla en la tabla 43.

Tabla 43.
Características de los dispositivos empleados.

Dispositivos	Tamaño de pantalla en pulgadas	Sistema operativo	Navegador de internet	Versión de navegador
Computador 1	13	MacOS Sierra	Safari	10.0.2
Computador 2	15	Windows 7	Mozilla	51.0.1
Tablet	7.9	IOS 2.2	Safari	10.0.2
Smartphone	4.7	Android 6.0.1	Chrome	56.0.29

El contenido de la página web se adapta a cada navegador de internet, tamaño de pantalla y sistema operativo permitiendo la correcta visualización, debido al uso de

código HTML5 permitiendo que la aplicación sea *responsive* para todos los dispositivos en los diferentes tamaños de pantalla, las pruebas se realizaron en el servidor local y externo sin presentar mayores variaciones. En las figuras 45,46,47,48,49,50,51 y 52 se visualiza las respectivas capturas de pantalla.



Figura 45. Visualización en computador1 en servidor local.



Figura 46. Visualización en computador2 en servidor local.



Figura 47. Visualización en tablet en servidor local.



Figura 48. Visualización en Smartphone en servidor local.



Figura 49. Visualización en computador 1 en servidor web.



Figura 50. Visualización en computador 2 en servidor web.



Figura 51. Visualización en tablet en servidor web.

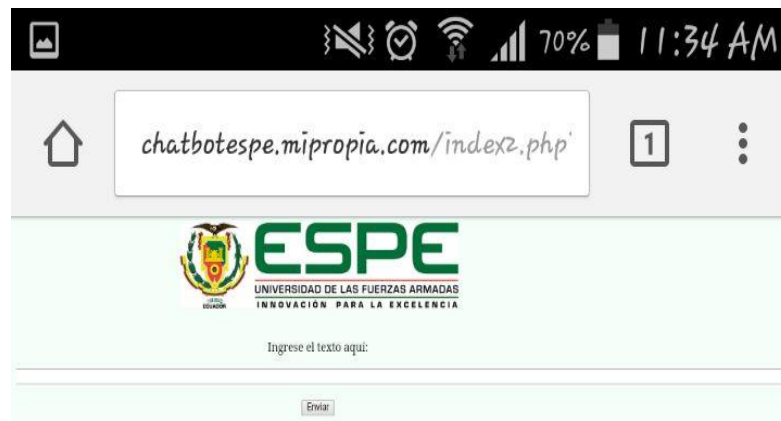


Figura 52. Visualización en Smartphone en servidor web.

1.1.23. Escenario de prueba 4: Usabilidad

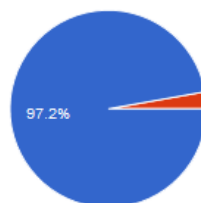
Para la realización de este escenario de prueba se realizó una encuesta de tipo “Formulario de Google”, con el fin de medir entre los usuarios: la facilidad de desplazamiento, el contenido de la base de datos, la respuesta del algoritmo y recomendaciones; las preguntas realizadas fueron las siguientes:

1. ¿Cómo es el acceso al aplicativo?
2. ¿Se visualiza correctamente en su dispositivo?
3. ¿La respuesta fue acorde con la pregunta realizada?
4. ¿Considera acorde la visualización de todo el historial de consultas realizadas?
5. ¿Existió ambigüedad en alguna respuesta obtenida?
6. ¿Considera usted completa la información al DEEE?
7. ¿Recomienda el uso de la aplicación?
8. ¿Qué considera que se debe mejorar?

1.1.23.1. Análisis de la facilidad de desplazamiento

Para este análisis se agruparon las preguntas: 1,2 y 4; con respecto a todo lo correspondiente a la visualización del aplicativo, donde se obtuvo los resultados que se muestran en las figuras 53,54 y 55:

¿Cómo es el acceso al aplicativo?

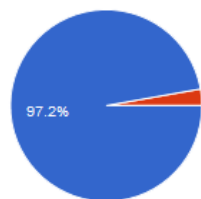


Facil	35	97.2%
Dificil	1	2.8%

Figura 53. Resultado de la pregunta 1.

El objetivo de esta pregunta es determinar la facilidad de acceso desde cualquier dispositivo al aplicativo al servidor externo, también se midió implícitamente la respuesta del *hosting* y con los resultados obtenidos se puede afirmar que cumplió con todas las expectativas necesarias para el desarrollo del proyecto.

¿Se visualiza correctamente en su dispositivo?



Si	35	97.2%
No	1	2.8%

Figura 54. Resultado de la pregunta 2.

Esta pregunta verificó el diseño web de la aplicación y su capacidad de visualización en todos los dispositivos. El 2.8% representa un error del despliegue del aplicativo, eso se debe al uso de navegadores antiguos o el uso de versiones obsoletas que no soportan HTML5 (*responsive*), como pueden ser: Internet Explorer en versiones inferiores a la 9; Safari en versiones inferiores a la 7 para dispositivos móviles e inferiores a versión 9 para escritorio.

¿Considera acorde la visualización de todo el historial de consultas realizadas?



Figura 55. Resultado de la pregunta 4.

El objetivo de la aplicación era almacenar y desplegar el historial que se realiza dentro del sistema a modo de chat, los resultados de esta pregunta despliegan que para el usuario no es necesario la visualización del historial debido al usos de los dispositivos móviles con pantallas reducidas generando incomodidad en el desplazamiento. Es por esto que para el servidor web se deshabilitó esta opción y se la mantiene el servidor local,

1.1.23.2. Análisis del contenido de la base de datos

El contenido de la base de datos se puso a prueba dentro de la pregunta 6, y los resultados obtenidos se muestran en la figura 56.

¿Considera usted completa la información correspondiente al DEEE?

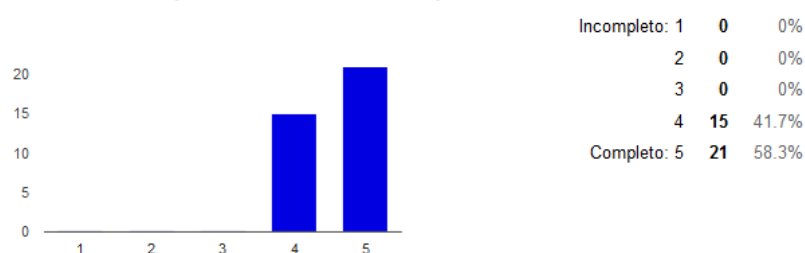


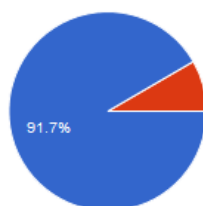
Figura 56. Resultado de la pregunta 6.

Dentro de la escala propuesta el resultado obtenido fue mayoritariamente positivo, es decir la base de datos contiene la información necesaria para el desarrollo de este aplicativo.

1.1.23.3. Análisis de la respuesta del algoritmo

En base a las pruebas realizadas en la sección 2.1 del capítulo 4, se obtuvo una eficiencia del algoritmo total superior al 92%, para corroborar este valor se realizó con las preguntas 3 y 5, resultados es mostrado en las figuras 57 y 58.

¿La respuesta fue acorde con la pregunta realizada?

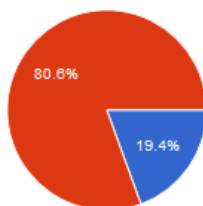


Si	33	91.7%
No	3	8.3%

Figura 57. Resultado de la pregunta 3.

El objetivo de esta pregunta se puede comparar con las preguntas realizadas en la sección 2.1 en el cual se busca determinar si el contenido de la respuesta tiene relación con la pregunta realizada, los errores que se pueden generar están asociados a posible falta de información en la base de datos sobre profesores antiguos que ya no pertenecen al DEEE, lugares de interés no considerados dentro del desarrollo del aplicativo, al igual que fechas no oficiales del calendario académico.

¿Existió ambigüedad en alguna respuesta obtenida?



Si	7	19.4%
No	29	80.6%

Figura 58. Resultado de la pregunta 5.

Esto se debe a que algunos profesores tienen el mismo apellido, lo que al momento de realizar el cálculo de índice de Jaccard despliega toda la información del apellido consultado más no de un docente específico. Del mismo caso se observó que para las materias que tienen secuencia se presenta el mismo inconveniente.

Se corrigió este error con un cambio en el algoritmo donde se presentó la ambigüedad con el aumento de un cálculo de Jaccard adicional con el que se buscó

filtrar el nombre del docente al igual que la materia específica. Este proceso se detalla en la figura 59 y 60.

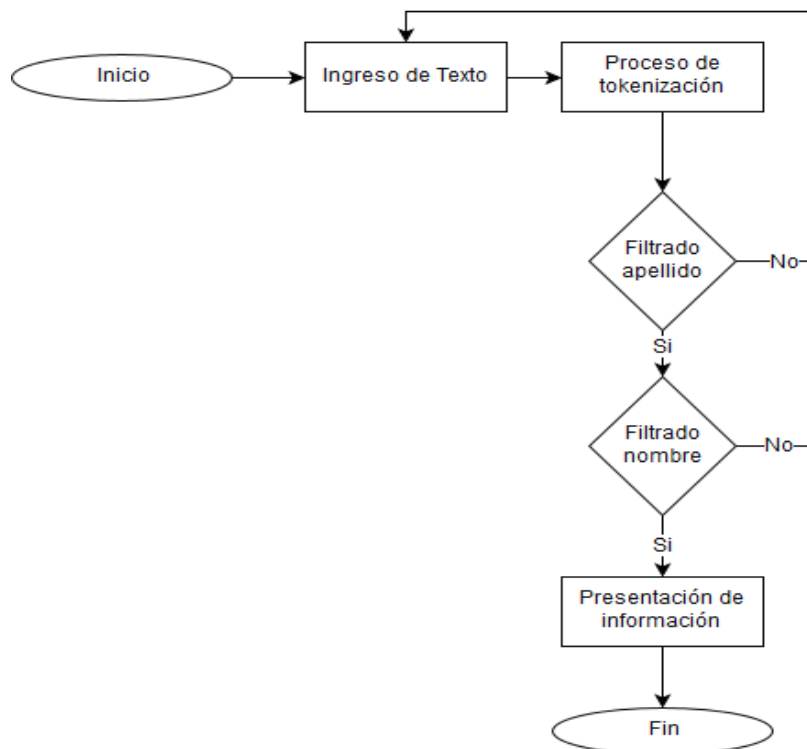


Figura 59. Reajuste de la tabla profesores.

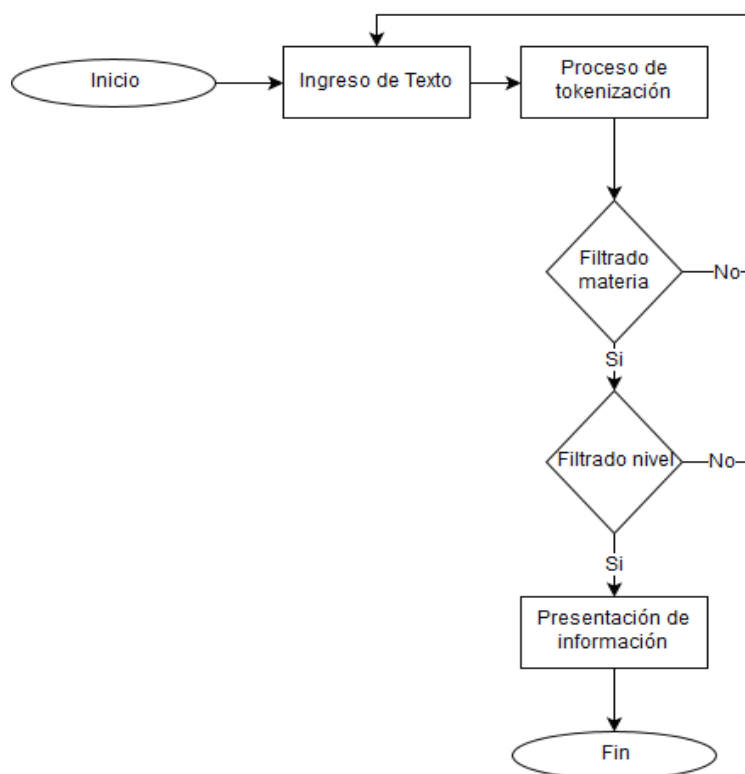


Figura 60. Reajuste de la tabla materias.

1.1.23.4. Análisis de recomendaciones

El análisis de las preguntas 7 y 8 consiste en verificar el uso de la aplicación y si el usuario está de acuerdo con lo realizado; al igual que observaciones breves generó el encuestado, en la figura 61.

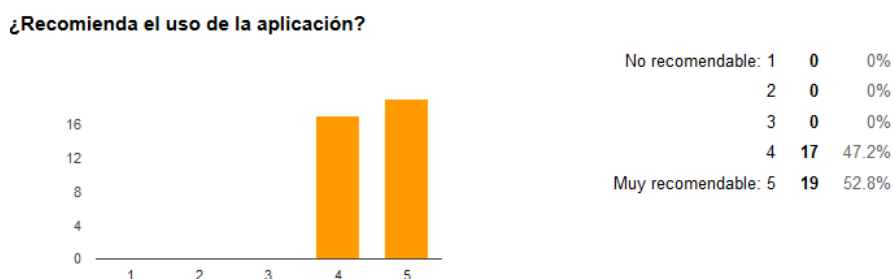


Figura 61. Resultado de la pregunta 7.

Los resultados de la recomendación de la aplicación son positivos, es decir el sistema fue aceptado por los usuarios al no recibir críticas negativas por lo cual se concluyó con el desarrollo. Con respecto a la pregunta 8 se pueden agrupar las

respuestas ingresadas en 2 grupos: el primero, indica respuesta positiva o que no se realice cambios al aplicativo; el segundo, indica la realización de mejoras como la ambigüedad y la eliminación del historial para mejor la visualización.

CAPÍTULO 5

5.1 Conclusiones

- En las pruebas sobre la eficiencia del algoritmo, se obtuvieron dos valores uno matemático y otro en base a encuestas realizadas a usuarios respectivamente, reflejando los siguientes porcentajes: 92.96%. este valor se obtuvo por medio de diferentes pruebas realizadas con preguntas ajustadas para el aplicativo y 91.70%, se obtuvo por medio de formularios de Google correspondientes a si las respuestas fueron acordes a las preguntas realizadas, después de realizar el reajuste en la ambigüedad en ciertas preguntas estos valores se aproximaron al 100%.
- El cálculo del índice de Jaccard evitó la creación de funciones específicas para el procesamiento de las frases tokenizadas ya que permitió un proceso de filtrado basado en un clasificador de palabras incluyendo espacios en blanco de manera ordenada, agilizando el acceso a la base de datos y permitiendo respuestas en tiempo real.
- En base a los datos obtenidos en la encuesta realizada en la pregunta sobre ambigüedad, se determinó que para eliminar este inconveniente, se agregó un índice de Jaccard como fue en el caso de los profesores en donde primero se filtró el apellido y luego el nombre, de igual manera este procedimiento se realizó con las materias de secuencia.
- La visualización del historial de consultas realizadas dentro del servidor web externo no se incluyó debido a problemas de conectividad al sitio FTP y su inestabilidad al momento de acceder a los datos almacenados en el archivo .txt esto fue corroborado en la encuesta ya que los usuarios no consideran que en sus dispositivos se despliegue todo el historial de consultas por lo que visualmente requiere de mucho espacio de desplazamiento en la pantalla de los dispositivos.
- Las pruebas realizadas sobre la visualización del aplicativo en los distintos dispositivos, sistemas operativos y navegadores fueron completamente satisfactorias mediante el uso de HTML5(*responsive*), para que esta característica funcione adecuadamente en todos los dispositivos estos deben estar actualizados en sistema operativo y navegador.

- El uso del Lenguaje Natural en la aplicación obligó a la eliminación de caracteres especiales (soportados por UTF-8) los cuales no son procesados por la herramienta PHP “php-nlp-tools” al igual que ciertos caracteres del español que generaron errores al momento de realizar las consultas a la base de datos por lo cual se mantuvo una codificación ASCII.

5.2 Recomendaciones

- Se recomienda el uso de servidores externos que permitan la instalación de ficheros adicionales como la librería de PLN “php-nlp-tools”, con la cual contiene las funciones utilizadas en este proyecto que es el cálculo del índice de Jaccard como fue el caso en el servidor *hosting* “mipropia.com”; debido a que este soporta el acceso FTP permitiendo almacenar archivos de una manera fluida.
- Se recomienda el uso del lenguaje de programación PHP para el desarrollo de aplicaciones web y manejo de base de datos como se realizó en este proyecto, debido a que este lenguaje facilita y agiliza el diseño de páginas gracias a su gran integrabilidad.
- Se recomienda la generación de encuestas periódicas como la desarrollada en este aplicativo para mantener actualizada la base de datos al igual que la eficiencia del algoritmo.

5.3 Trabajos futuros

- La base de datos implementada contiene actualmente información detallada del DEEE lo que implica que se puede desarrollar paulatinamente una ampliación que abarque a todos los departamentos de la Universidad de las Fuerzas Armadas-ESPE, al igual que complementar la información sobre los eventos actuales y futuros de la institución.
- Se propone realizar mejoras en el proceso de actualización de datos mediante la creación de dos usuarios: uno administrador y otro general

donde el usuario administrador mediante el uso del chat pueda actualizar los datos sin la necesidad de modificar información en la base de datos.

- Desplegar una red lan dentro de la Universidad para montar el servicio del chatbot en los principales lugares de confluencia de personas.
- Desarrollar una aplicación para Android-IOS que permita la interacción nativa desde el Smartphone.

REFERENCIAS

- Aetecno. (2016). *¿Qué son los Bots y por qué llegarán en un futuro a todas las conversaciones?* Obtenido de <http://tecno.americaeconomia.com/articulos/que-son-los-bots-y-por-que-llegaran-en-un-futuro-todas-tus-conversaciones>
- Amazon Inc. (2015). *Amazon Echo*. Obtenido de <https://www.amazon.com/Amazon-Echo-Bluetooth-Speaker-with-WiFi-Alexa/dp/B00X4WHP5E>
- Amazon Mechanical Turk. (Noviembre de 2015). *Amazon Mechanical Turk*. Obtenido de <http://www.mturk.com>
- Apr. (2016). *¿Cuáles son las versiones de HTML? Diferencias entre HTML 4 y HTML 5.* Obtenido de http://aprenderaprogramar.com/index.php?option=com_content&view=article&id=444:icuales-son-las-versiones-de-html-diferencias-entre-html-4-y-html-5-significado-de-strict-cu00706b&catid=69:tutorial-basico-programador-web-html-desde-cero&Itemid=192
- Aref, B. (2005). *A Corpus Based Approach to Generalising a Chatbot System*. University of Leeds, Leeds. Obtenido de <http://etheses.whiterose.ac.uk/1323/1/abushawar.pdf>
- Baeza, O. (2005). *XAMPP: Servidor Web, Correo, Base de Datos y más Extras en Pocos Minutos*. Obtenido de <http://www.genbeta.com/web/xampp-servidor-web-correo-base-de-datos-y-mas-extras-en-pocos-minutos>
- Best-Rowden, L. B. (Septiembre 2014). Unconstrained face recognition: Establishing baseline human performance via crowdsourcing. *In Biometrics (IJCB), 2014 IEEE International Joint Conference on . IEEE.*, (pp. 1-8).
- BioSecure, A. (2015). *Bio Secure*. Obtenido de <http://biosecure.it-sudparis.eu/>
- Blázquez, M. (Diciembre de 2012). *Técnicas Avanzadas de Recuperación de Información*. Obtenido de <http://ccdoc-tecnicasrecuperacioninformacion.blogspot.com/2012/12/modelo-vectorial.html>

- Bohn, D. (2016). *Google Home: a speaker to finally take on the Amazon Echo*. Obtenido de <http://www.theverge.com/2016/5/18/11688376/google-home-speaker-announced-virtual-assistant-io-2016>
- Buitelaar, P. (2014). *Natural Language Processing- for the Semantic Web* -. National University Of Irland, DERI. Obtenido de <https://www.deri.ie/fileadmin/documents/teaching/tutorials/DERI-Tutorial-NLP.final.pdf>
- Bykarskij, E. (2015). *Microsoft Cortana*. Obtenido de <http://oc.sti2.at/sites/default/files/cortana.pdf>
- Casillas, A. (2013). *Bases de Datos*. Obtenido de http://ocw.uoc.edu/computer-science-technology-and-multimedia/bases-de-datos/bases-de-datos/P06_M2109_02151.pdf
- Castillo, A. (2011). *Diseño y Desarrollo de Base de Datos en MySQL y Aplicación Web en PHP con Servidor Central APACHE*. Universidad Politécnica de Valencia, Informática, Valencia. Obtenido de <https://riunet.upv.es/bitstream/handle/10251/11166/memoria.pdf?sequence=1>
- Chao, A. (2004). *Un Nuevo Método Estadístico Para la Evaluación de la Similitud en la Composición de Especies con Datos de Incidencia y Abundancia*. Ecology Letters. Obtenido de <http://viceroy.eeb.uconn.edu/estimates/EstimateSPages/EstSUsersGuide/References/ChaoEtAl2005Sp.pdf>
- Cobos, J. C. (2013). *Integración de un Chatbot Como Habilidad de un Robot Social con Gestor de Diálogos*. Ingeniería De Sistemas Y Automática, Madrid. Obtenido de <http://repositorio.educacionsuperior.gob.ec/bitstream/28000/1201/1/T-SENESCYT-000332.pdf>
- Cortez, A., Vega, H., & Pariona, J. (2013). *Procesamiento del Lenguaje Natural*. Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática.
- David, E. R. (2005). *Prototipo de software de computadora para reconocimiento de firmas manuscritas*.

- (s.f.). *Desarrollo e implementación de un Bot Conversacional Como Apoyo a los Estudiantes en su Proceso de titulación.*
- Díaz, M. M. (2008). Dynamic signature verification for portable devices.
- Díaz-Cabrera, M. F. (2015). Modeling the Lexical Morphology of Western Handwritten Signatures.
- Eguíluz, J. (2008). *Introducción a JavaScript*. Obtenido de <https://adegiusti.files.wordpress.com/2013/09/el-gran-libro-de-html5-css3-y-javascript.pdf>
- Enríquez, A., & Maldonado, J. (2010). *MySQL*. UAEM. Obtenido de <http://www.gridmorelos.uaem.mx/~mcruz/cursos/miic/MySQL.pdf>
- Estellés-Arolas, E. &.-L.-d.-G. (2012). Towards an integrated crowdsourcing definition. *Journal of Information science*, 189-200.
- Faundez-Zanuy, M. .. (2005). Signature recognition state-of-the-art. ., *Aerospace and Electronic Systems Magazine, IEEE*, 28-32.
- Fierrez, J. G.-G.-F. (2010). BiosecuID: a multimodal biometric database. *Pattern Analysis and Applications*. 235-246.
- García, M., & Valle, L. (2013). *Desarrollo e Implementación de un Bot Conversacional Como Apoyo a los Estudiantes en su Proceso de Titulación.* Universidad Autónoma de Yucatán.
- Gauchat, J. D. (2012). *El Gran Libro de HTML 5, CS3 y JavaScript*. Marcombo. Obtenido de <https://adegiusti.files.wordpress.com/2013/09/el-gran-libro-de-html5-css3-y-javascript.pdf>
- González, E. (2016). *¿Es PHP un buen lenguaje de programación? ¿Cuáles son las versiones de PHP? Historia.* Obtenido de http://aprenderaprogramar.com/index.php?option=com_content&view=article&id=512:ies-php-un-buen-lenguaje-de-programacion-icuales-son-las-versiones-de-php-historia-cu00806b&catid=70:tutorial-basico-programador-web-php-desde-cero&Itemid=193
- Guru, D. S. (2009). Online signature verification and recognition: An approach based on symbolic representation. *Pattern Analysis and Machine Intelligence. IEEE Transactions on*, 1059-1073.

- Harrison, D. B. (2009). Handwriting examination: Meeting the challenges of science and the law. *Forensic Science Communications*.
- Hunter, K. (2015). *Cortana and Speech Platform Overview*. Obtenido de <https://sec.ch9.ms/slides/winHEC/CortanaandSpeechPlatformOverview.pdf>
- Ibiblio. (2014). *Curso completo de HTML*. Obtenido de <http://ibiblio.org/pub/linux/docs/LuCaS/Manuales-LuCAS/doc-curso-html/doc-curso-html/x38.html>
- Ipeirotis, P. G. (2010). Demographics of mechanical turk.
- Ituarte, L. E. (2006). Hacia una evaluación cuantitativa de los detectores de esquinas: el caso de las curvas ROC empíricas. *XII Congreso Argentino de Ciencias de la Computación*.
- Jibo Inc . (2016). *Meet Jibo, The World's First Social Robot*. Obtenido de <https://www.jibo.com>
- Jibo Inc. (2016). *The Jibo SDK: Reaching out Beyond the Screen*. Obtenido de <https://developers.jibo.com/blog/the-jibo-sdk-reaching-out-beyond-the-screen#>
- Jiménez , P., & Sánchez , J. (2015). *De Eliza a Siri: La Evolución*. Obtenido de <http://www.uax.es/publicacion/de-eliza-a-siri-la-evolucion.pdf>
- Junta de Andalucía. (2010). *Apache Web Server*.
- kabir, M. (2006). *La Biblia del Servidor Apache*.
- Ken, D. (2015). *Características, Ventajas y Desventajas de las Tecnologías de Desarrollo Ruby on Rails y PHP*. Obtenido de <https://jditic92.wordpress.com/2015/05/11/caracteristicas-ventajas-y-desventajas-de-las-tecnologias-de-desarrollo-ruby-on-rails-y-php/>
- Mateos, M., & Reina, R. (2013). *Procesamiento del Lenguaje Natural*. Universidad de Sevilla, Ciencias de la Computación e Inteligencia Artificial, Sevilla. Obtenido de <https://www.cs.us.es/cursos/ia2/temas/tema-06.pdf>
- MDN. (Mayo de 2016). *Guía del desarrollador HTML*. Obtenido de <https://developer.mozilla.org/es/docs/Web/Guide/HTML>
- Neoteo. (Octubre de 2012). *10 útiles Chatbots Para Mensajería Instantánea*. Obtenido de <http://www.neoteo.com/10-utiles-chatbots-para-mensajeria-instantanea>

- Oracle. (2011). *MySQL 5.0 Reference Manual*. Obtenido de <http://downloads.mysql.com/docs/refman-5.0-es.pdf>
- Palomo, M., & Montero, I. (2011). *Programación en PHP a través de Ejemplos*. Universidad de Cádiz, Cádiz. Obtenido de http://servicio.uca.es/softwarelibre/publicaciones/apuntes_php
- Pavón, J. (2013). *Aplicaciones Web/Sistemas Web-Servidor Web Apache*. Universidad Complutense Madrid, Facultad de Informática, Madrid.
- Paz, M. (Diciembre de 2012). *6 buenos motivos para trabajar con PHP*. Obtenido de <https://www.lancetalent.com/blog/6-buenos-motivos-para-trabajar-con-php/>
- Pazos, A., & Barreiro, J. (1999). *Teleinformática*.
- Peinado, F. (2010). *PHP Introducción y Sintáxis*. Universidad Complutense de Madrid, Madrid. Obtenido de <https://www.fdi.ucm.es/profesor/fpeinado/courses/webtech/Tema7-PHP.pdf>
- PHP. (2015). *NlpTools*. Obtenido de <http://php-nlp-tools.com/documentation/>
- PHP. (2015). *NlpTools API*. Obtenido de <http://php-nlp-tools.com/documentation/api/>
- PHP. (9 de junio de 2016). *php*. Obtenido de <http://php.net/>
- Polatidis, N. (2011). *Chatbot for Admissions*. University of Birmingham, Birmingham. Obtenido de <https://arxiv.org/pdf/1408.6762.pdf>
- President's Council of Advisors on Science and Technology. (2010). *Apple's Siri: A Closer Look*. Obtenido de <http://cs.fit.edu/~pkc/dept/siri/siriIntro.pdf>
- Programando la web. (2016). *Ventajas y Desventajas en el Uso de PHP*. Obtenido de <http://www.programandolaweb.com/posts/php/6/Ventajas-y-desventajas-en-el-uso-de-php.html>
- Quentin. (2015). *The Ultimate Guide To Mastering Amazon Echo*. Obtenido de <http://blog.anythings.co/ultimate-guide-amazon-echo-alexa/>
- Ramírez, Rubén;. (2011). *Un Robot conversador emotivo*. Barcelona. Obtenido de <https://upcommons.upc.edu/bitstream/handle/2099.1/13801/76688.pdf?sequence=1&isAllowed=y>
- Rodríguez Correa, S., & Benavides, P. (2007). *Procesamiento de Lenguaje Natural en la Recuperación de información*. Univeridad de la Salle.
- Russell, S., & Norvig, P. (2004). *Inteligencia Artificial: Un Enfoque Moderno* (Segunda ed.). Madrid: Pearson.

- Sánchez, J. (2013). *Manual de Referencia de JavaScript*. Obtenido de <http://www.jorgesanchez.net/web/javascript.pdf>
- Santa María, L. (2014). *12 Ventajas de la Programación PHP que Debes Saber*. Obtenido de <http://www.staffcreativa.pe/blog/ventajas-programacion-php/>
- Shrager, J. (2015). *The Genealogy of Eliza*. Obtenido de <http://elizagen.org/index.html>
- Styde. (Octubre de 2014). *Qué es Composer y cómo usarlo*. Obtenido de <https://styde.net/que-es-composer-y-como-usarlo/>
- Toscano Medina, L. K. (2009). *RECONOCIMIENTO DINAMICO Y ESTATICO DE TRAZOS*. Doctoral Dissertation.
- Turing, A. (1950). *Computing Machinery and Intelligence*. Mind. Obtenido de <http://www.csee.umbc.edu/courses/471/papers/turing.pdf>
- Valle, L., & García, M. (2013). *Desarrollo e Implementación de un Bot Conversacional Como Apoyo a los Estudiantes en su Proceso de Titulación*. Universidad Autónoma de Yucatán, Facultad de Matemáticas.
- Valle, L., García, M., & López, J. (2013). *Desarrollo e Implementación de un Bot Conversacional Como Apoyo a los Estudiantes en su Proceso de Titulación*. Universidad Autónoma de Yucatán, Facultad de Matemáticas, Unidad Multidisciplinaria Tizimín.
- Vallez, M. (2007). *El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y Áreas Afines*. Obtenido de <https://www.upf.edu/hipertextnet/numero-5/pln.html>
- Van Lun, E. (2014). *Chatbots By BotGenes*. Obtenido de <https://www.chatbots.org/developer/botgenes/>
- Vázquez, C. (2008). *Programación en PHP5. Nivel Básico*. Obtenido de http://www.ticarte.com/sites/su/users/7/arch/manual_php5_basico.pdf
- Vela, A. (2015). *Clasificación de Bots y Trolls #INFORGAFIA #INFOGRAPHIC*. Obtenido de <https://ticsyformacion.com/2015/09/07/clasificacion-de-bots-y-trolls-inforgafia-infographic/>
- Velez, J. F. (2009). Snake models for offline signature verification: a comparative study. *International Journal of Innovative Computing and Applications*, (págs. 34-44.).

- Vílchez, E. (2007). *Elementos Básicos de Programación con PHP*. Universidad Nacional de Costa Rica, Escuela de Informática, San José. Obtenido de <http://www.cidse.itcr.ac.cr/ciemac/memorias/5toCIEMAC/Talleres/ElementosbasicosdeprogramacionconPHP.pdf>
- Web Docente Departamental. (2010). *Los Lenguajes de Marcas*. Obtenido de http://www.deciencias.net/disenoweb/elaborardw/paginas/intro_html.htm
- Webadictos. (Diciembre de 2012). *Breve Historia del HTML*. Obtenido de [https://webadictos.com/2012/12/30/breve-historia-del-html/Breve Historia del HTML](https://webadictos.com/2012/12/30/breve-historia-del-html/Breve%20Historia%20del%20HTML)
- Winarsky, N., Mark, B., & Kressel, H. (2013). *The development of Siri and the SRI Venture Creation Process*. Obtenido de http://www.mauldineconomics.com/images/uploads/overmyshoulder/Siri_Caltech_Case_HK_-_Bill_Norman_v11_5.pdf