

**UNIVERSIDAD PERUANA UNIÓN**  
FACULTAD DE INGENIERÍA Y ARQUITECTURA  
Escuela Profesional de Ingeniería de Sistemas



*Una Institución Adventista*

**Motor de Búsqueda basado en el Modelo de Espacio Vectorial  
para mejorar la difusión de la información científica de las  
especies de mamíferos inventariados en la Región San Martín.**

Por:

Gerson Kieffer Layza Guerrero

Asesor:

Mg. Pedro Antonio Gonzales Sánchez

**Tarapoto, noviembre de 2019**

## DECLARACIÓN JURADA DE AUTORIA DEL INFORME DE TESIS

*Mg. Pedro Antonio Gonzales Sánchez*, de la Facultad de Ingeniería y Arquitectura, Escuela Profesional de Ingeniería de Sistemas, de la Universidad Peruana Unión.

### DECLARO:

Que el presente informe de investigación titulado: ***“Motor de Búsqueda basado en el Modelo de Espacio Vectorial para mejorar la difusión de la información científica de las especies de mamíferos inventariados en la Región San Martín”***, constituye la memoria que presenta el **Bachiller Layza Guerrero, Gerson Kieffer**; para aspirar al título Profesional de Ingeniero de Sistemas, que ha sido realizada en la Universidad Peruana Unión, bajo mi dirección.

Las opiniones y declaraciones en este informe son de entera responsabilidad del autor, sin comprometer a la institución.

Y estando de acuerdo, firmo la presente constancia en Morales, a los 04 días del mes de noviembre del año 2019.



---

Mg. Pedro Antonio Gonzales Sánchez

**Asesor**

**Motor de Búsqueda basado en el Modelo de Espacio Vectorial  
para mejorar la difusión de la información científica de las  
especies de mamíferos inventariados en la Región San Martín.**

## **TESIS**

Presentada para optar el título profesional de Ingeniero de Sistemas

### **JURADO CALIFICADOR**

  
Mg. Danny Lévano Rodríguez  
Presidente

  
Mg. Joseph Ibrahim Cruz Rodríguez  
Secretario

  
Mg. Immer Elías Cuellar Rodríguez  
Vocal

  
Mg. Pedro Antonio Gonzales Sanchez  
Asesor

**Morales, 04 de noviembre del año 2019**

## **DEDICATORIA**

A mis queridos padres, por su apoyo incondicional, a mi hermana, amigos y seres queridos que fueron la inspiración para lograr este objetivo.

## **AGRADECIMIENTO**

A Dios en primer lugar por darme la vida, guiarme siempre cada día y darme la sabiduría para alcanzar mis metas.

A mis padres por su amor incondicional y apoyo abnegado en la parte económica y espiritual.

A la Universidad Peruana Unión sede Tarapoto, por su hospitalidad académica durante estos 5 años, por su enseñanza pedagógica y en especial por su experiencia espiritual que legitimaron mi formación profesional.

## Índice general

Índice general.....	vi
Índice de tablas .....	x
Índice de figura .....	xi
Índice de gráficos.....	xiii
Índice de anexos.....	xiv
Lista de símbolos y términos .....	xv
RESUMEN .....	xvi
ABSTRACT.....	xvii
Capítulo 1 EL PROBLEMA.....	18
Capítulo 2 REVISIÓN DE LA LITERATURA .....	23
2.1. Recuperación de información .....	23
2.2. Recuperación de la información en la world wide web.....	24
2.2.1. Recuperación en la Web .....	24
2.2.2. Técnicas de recuperación en la Web .....	24
2.3. Modelos de recuperación de información.....	25
2.3.1. Modelo Booleano.....	25
2.3.2. Modelo Vectorial .....	25
2.3.3. Modelo Probabilístico.....	26
2.4. Modelo de Espacio Vectorial (MEV) .....	27
2.4.1. Factor Term Frequency (TF) .....	28
2.4.2. Factor Inverse Document Frequency (IDF).....	30
2.4.3. Ponderación TF-IDF.....	31
2.5. Motor de búsqueda .....	32
2.6. Difusión de la información .....	33
2.6.1. Formas de difusión .....	33
2.6.2. Difusión bajo demanda .....	33
2.6.3. Difusión documental.....	33
2.7. Biodiversidad .....	35
2.8. Minería de datos .....	35
2.9. Modelo de procesos para proyectos de minería de datos .....	36
2.9.1. Modelo CRISP-DM.....	36
Capítulo 3 MATERIALES Y MÉTODOS .....	44
3.1. Tipo de investigación.....	44

3.2.	Descripción del lugar de ejecución .....	44
3.3.	Población y muestra.....	45
3.3.1.	Población .....	45
3.3.2.	Muestra .....	45
3.4.	Diseño de investigación.....	46
3.4.1.	Analizar.....	47
3.4.2.	Evaluar pretest .....	47
3.4.3.	Implementar .....	47
3.4.4.	Evaluar postest.....	47
3.4.5.	Validar .....	48
3.5.	Formulación de la hipótesis .....	48
3.5.1.	Hipótesis principal .....	48
3.5.2.	Hipótesis específicas.....	48
3.6.	Identificación de variables.....	48
3.6.1.	Matriz de consistencia .....	49
3.6.2.	Operacionalización de variables .....	51
3.7.	Instrumento de recolección de datos.....	53
3.8.	Técnica de recolección de datos .....	53
3.9.	Validación de instrumentos .....	53
3.10.	Plan de procesamiento de datos .....	53
Capítulo 4	CONSTRUCCIÓN DE LA PROPUESTA .....	54
4.1.	Introducción.....	54
4.2.	Analizar.....	54
4.2.1.	Generalidades de la empresa .....	54
4.3.	Evaluar pretest .....	58
4.4.	Implementar .....	60
4.4.1.	Comprensión de los datos .....	60
4.4.2.	Modelado de los datos .....	62
4.4.3.	Evaluación .....	63
4.4.4.	Implantación .....	67
4.4.5.	Realizar Consulta.....	67
4.4.6.	Representación de la colección de documentos.....	69
4.4.7.	Representación de una lista invertida .....	69
4.4.8.	Cálculo del IDF a partir de la lista invertida.....	70

4.4.9.	Ponderación TF – IDF .....	71
4.4.10.	Cálculo de similaridad del documento con la consulta .....	73
4.4.11.	Ranking de consulta.....	74
4.5.	Evaluar postest.....	76
4.6.	Validar .....	77
4.6.1.	Formulación de la hipótesis estadística: .....	77
4.6.2.	Nivel de significancia .....	77
4.6.3.	Estadístico de prueba .....	77
4.6.4.	Determinación de la región crítica:.....	78
4.6.5.	Obtención del valor experimental:.....	78
4.6.6.	Decisión .....	81
4.6.7.	Conclusión .....	81
Capítulo 5	RESULTADOS Y DISCUSIÓN .....	82
5.1.	Difusión bajo demanda .....	82
5.2.	Difusión documentaria.....	84
5.3.	Medio de difusión utilizado .....	86
5.4.	Sobre el Buscador Informático .....	88
5.4.1.	Nivel de similitud .....	88
5.4.2.	Navegación intuitiva.....	90
5.4.3.	Interfaz del buscador.....	91
5.5.	Influencia de la variable X sobre la variable Y .....	92
Capítulo 6	CONCLUSIONES Y RECOMENDACIONES .....	93
6.1.	Conclusiones.....	93
6.2.	Recomendaciones .....	94
Capítulo 7	REFERENCIAS BIBLIOGRAFICAS.....	95
Capítulo 8	ANEXOS .....	99
8.1.	Pretest .....	99
8.2.	Postest.....	101
8.3.	Preguntas de entrevista .....	104
8.4.	Validación de juicio de expertos.....	105
8.5.	Detalle del desarrollo del motor de búsqueda.....	108
8.5.1.	Funcionalidades del negocio.....	108
8.5.2.	Modelado del negocio.....	108
8.5.3.	Diagrama de despliegue.....	116



8.5.4.	Seudocódigo del algoritmo de lista invertida .....	117
8.5.5.	Seudocódigo del algoritmo de búsqueda .....	119
8.6.	Pantallazos del sistema .....	121
8.7.	Manual del usuario .....	123
8.7.1.	Introducción .....	125
8.7.2.	Requerimientos técnicos .....	126
8.7.3.	Herramientas utilizadas para el desarrollo .....	126
8.7.4.	Funcionalidad principal .....	127
8.7.5.	Funcionalidad de administrador .....	129
8.7.6.	Interfaz del administrador .....	130
8.7.7.	Módulos que Integran el sistema .....	135

## Índice de tablas

Tabla 1	Características del Factor TF .....	29
Tabla 2	Características del factor IDF .....	30
Tabla 3	Matriz de consistencia .....	49
Tabla 4	Operacionalización de variables .....	51
Tabla 5	Objetivos estratégicos institucionales .....	55
Tabla 6	Representación final del vocabulario .....	70
Tabla 7	Lista invertida .....	71
Tabla 8	Ponderación de los términos .....	72
Tabla 9	Cálculo de Similaridad del documento con la consulta .....	73
Tabla 10	Ranking de la consulta .....	75
Tabla 11	Resumen de datos calculados .....	81
Tabla 12	Pretest difusión bajo demanda .....	82
Tabla 13	Postest difusión bajo demanda .....	83
Tabla 14	Pretest difusión documentaria .....	85
Tabla 15	Postest difusión documentaria .....	85
Tabla 16	Pretest medio de difusión utilizado .....	86
Tabla 17	Postest medio de difusión utilizado .....	87
Tabla 18	Postest nivel de similitud .....	89
Tabla 19	Postest navegación intuitiva .....	90
Tabla 20	Postest interfaz del navegador .....	91
Tabla 21	Consulta de inventario de mamíferos .....	109
Tabla 22	Mantenimiento de inventarios de mamíferos .....	110
Tabla 23	Seguridad .....	110
Tabla 24	ECUR Seguridad .....	114
Tabla 25	ECUR Mantenimiento: Registrar especie .....	115
Tabla 26	ECUR Mantenimiento: Generar lista invertida .....	116
Tabla 27	Botones principales .....	132

## Índice de figura

Figura 1 Total de registros por zonas en el Perú.....	19
Figura 2 Modelo de relevancia probabilístico .....	27
Figura 3 Buscadores más populares de la historia .....	28
Figura 4 Cálculo de la frecuencia de aparición.....	29
Figura 5 Cálculo del IDF de un término .....	31
Figura 6 Ponderación TF-IDF.....	31
Figura 7 Search Engine Market Share Worldwide Mar 2017 – Mar 2018.....	32
Figura 8 Metodologías más utilizadas del Data Mining .....	37
Figura 9 Fase de comprensión del negocio.....	38
Figura 10 Fase de comprensión de los datos. ....	39
Figura 11 Fase de preparación de los datos .....	40
Figura 12 Fase de modelado .....	41
Figura 13 Fase de evaluación.....	42
Figura 14 Fase de implantación .....	43
Figura 15 Diseño de investigación.....	47
Figura 16 Preguntas de entrevista parte 1 .....	56
Figura 17 Preguntas de entrevista parte 2 .....	57
Figura 18 Aplicando encuesta algunos representantes .....	58
Figura 19 Pretest online de Google.....	59
Figura 20 Base de datos en Excel .....	61
Figura 21 Base de datos del sistema .....	62
Figura 22 Evaluando resultados del MB -1 .....	63
Figura 23 Evaluando resultados del MB -2 .....	64
Figura 24 Tiempo requerido para realizar búsqueda .....	64
Figura 25 Evaluando la función editar .....	65
Figura 26 Evaluando la función agregar .....	66
Figura 27 Evaluando la función eliminar.....	66
Figura 28 Consulta de un usuario .....	68
Figura 29 Representación de la colección de documentos .....	69
Figura 30 Resultados de la consulta "Mono" .....	75
Figura 31 Postest.....	76
Figura 32 Región crítica.....	78
Figura 33 Diagrama de caso de uso de negocio.....	109
Figura 34 MON Búsqueda de inventarios. ....	111
Figura 35 MON Mantenimiento .....	112
Figura 36 MON Seguridad.....	112
Figura 37 MCUR Seguridad. ....	113
Figura 38 MCUR mantenimiento. ....	113
Figura 39 MUCR búsqueda de mamíferos .....	114
Figura 40 Diagrama de despliegue. ....	116
Figura 41 Página principal del buscador.....	127
Figura 42 Ingreso valores para búsqueda.....	128
Figura 43 Ejecución de búsqueda .....	128
Figura 44 Resultados de búsqueda.....	129
Figura 45 Ingreso al módulo administrador .....	130
Figura 46 Inicio de sesión al módulo administrador.....	130
Figura 47 Modulo del administrador .....	131

Figura 48 Botones principales .....	132
Figura 49 Botón Guardar .....	133
Figura 50 Botón editar .....	133
Figura 51 Botón eliminar .....	133
Figura 52 Botón agregar .....	134
Figura 53 Vista principal - Modulo administrador .....	135
Figura 54 Taxones - reino .....	136
Figura 55 Taxones - filos .....	136
Figura 56 Taxones - clases.....	137
Figura 57 Taxones - ordenes.....	137
Figura 58 Taxones - familias .....	138
Figura 59 Taxones – géneros .....	138
Figura 60 Taxones - especies de mamíferos .....	139
Figura 61 Módulo de seguridad .....	139
Figura 62 Módulo de lista invertida.....	140

## Índice de gráficos

Gráfico 1 Pretest difusión bajo demanda .....	83
Gráfico 2 Postest difusión bajo demanda.....	84
Gráfico 3 Pretest difusión documentaria.....	85
Gráfico 4 Postest difusión documentaria .....	86
Gráfico 5 Pretest medio de difusión utilizado.....	87
Gráfico 6 Postest medio de difusión utilizado .....	88
Gráfico 7 Postest nivel de similitud .....	89
Gráfico 8 Postest navegación intuitiva .....	90
Gráfico 9 Postest interfaz del navegador .....	91
Gráfico 10 Influencia de x en y.....	92

## Índice de anexos

Anexos 1 Pre Test - P1.....	99
Anexos 2 Pre Test - P2.....	100
Anexos 3 Post Test - P1 .....	101
Anexos 4 Post Test - P2 .....	102
Anexos 5 Post Test - P3 .....	103
Anexos 6 Preguntas de entrevistas.....	104
Anexos 7 Validación del Juez n° 1 .....	105
Anexos 8 Validación del Juez n° 2 .....	106
Anexos 9 Validación del Juez n° 3 .....	107
Anexos 10 Página principal del motor de búsqueda de mamíferos .....	121
Anexos 11 Resultados de búsqueda .....	121
Anexos 12 Panel de control del motor de búsqueda .....	122
Anexos 13 Inventario de especie de mamíferos .....	122

## Lista de símbolos y términos

<b>TF</b>	Frecuencia del Término
<b>IDF</b>	Frecuencia Inversa del Documento
<b>FLORA</b>	Conjunto de plantas de una determinada zona.
<b>FAUNA</b>	Conjunto de especies de animales con referencia de un lugar, clima, tipo, medio o periodo geológico.

## RESUMEN

La presente investigación tiene como finalidad mejorar la difusión de la información científica de mamíferos inventariados en la región San Martín, a través de un motor de búsqueda basado en el modelo de espacio vectorial. Parte de la elaboración del buscador, consistió en realizar una serie de procedimientos establecidos en nuestra metodología, con las siguientes fases: análisis, evaluación pretest, implementación, evaluación posttest, y validación. El análisis situacional tuvo lugar en el Instituto de Investigaciones de la Amazonía Peruana de San Martín, se realizó entrevistas con respecto a la información científica de fauna con la que se cuenta en la región. Después, seleccionamos 12 representantes de instituciones comprometidas con la conservación de la biodiversidad en San Martín, a los cuales se aplicó instrumentos para medir la percepción en cuanto a la calidad de difusión de la información científica, antes y después la implementación.

Por otro lado, para el desarrollo del motor de búsqueda se utilizó el lenguaje de programación PHP, y PostgreSQL para la gestión de la base de datos relacional. Así mismo contempla las siguientes funcionalidades en su administración interna: registro de la clasificación científica (taxón) como especies de mamíferos, género, familia, ordenes, clases, filos y reinos, además cuenta con el buscador que es la interfaz principal para el usuario.

Los resultados de la investigación muestran que, del total de encuestados en el pre test, respecto al medio de difusión, el 42% de representantes lo califica como malo, y un 50% lo califica como regular, por lo cual concluimos que el medio de difusión es regular; en cuanto al post test, el 25% de los representantes lo califica como muy bueno, y un 58% lo califica como bueno, esto confirma que el uso del motor de búsqueda como medio de difusión de la información científica de mamíferos, logró una excelente aprobación.

Palabras claves: motor de búsqueda, modelo de espacio vectorial.



## **ABSTRACT**

This study aims to improve the dissemination of scientific information regarding mammals inventoried in the San Martin region, through a search engine based on the vector space model. Part of the elaboration of the search engine consisted in carrying out a series of procedures established in our methodology, with the following phases: analysis, pre-test evaluation, implementation, post-test evaluation, and validation. The situational analysis took place at the Peruvian Amazon Research Institute of San Martin, and interviews were conducted with respect to the scientific information available for fauna in the region. Then, we select 12 representatives of institutions committed to the conservation of biodiversity in San Martin, to whom instruments were applied to measure their perception regarding the quality of dissemination of scientific information. Before and after the implementation.

As for the development of the search engine, the programming language PHP was used, and PostgreSQL for the management of the relational database. Likewise, contemplates the following functionalities in its internal administration: registration of the scientific classification (taxon) for mammal species, genus, family, orders, classes, phylums and kingdoms, in addition to the search engine which is the main interface for the user.

The results from this study show that, out of the total number of people surveyed in the pre-test, with respect to the media, 42% of the representatives qualify it as bad, and 50% qualify it as regular, for which we conclude that the media is regular; as for the post-test, 25% of the representatives qualify it as very good, and 58% qualify it as good, this confirms that the use of the search engine as a means of disseminating scientific information regarding mammals, achieved an improvement in perception.

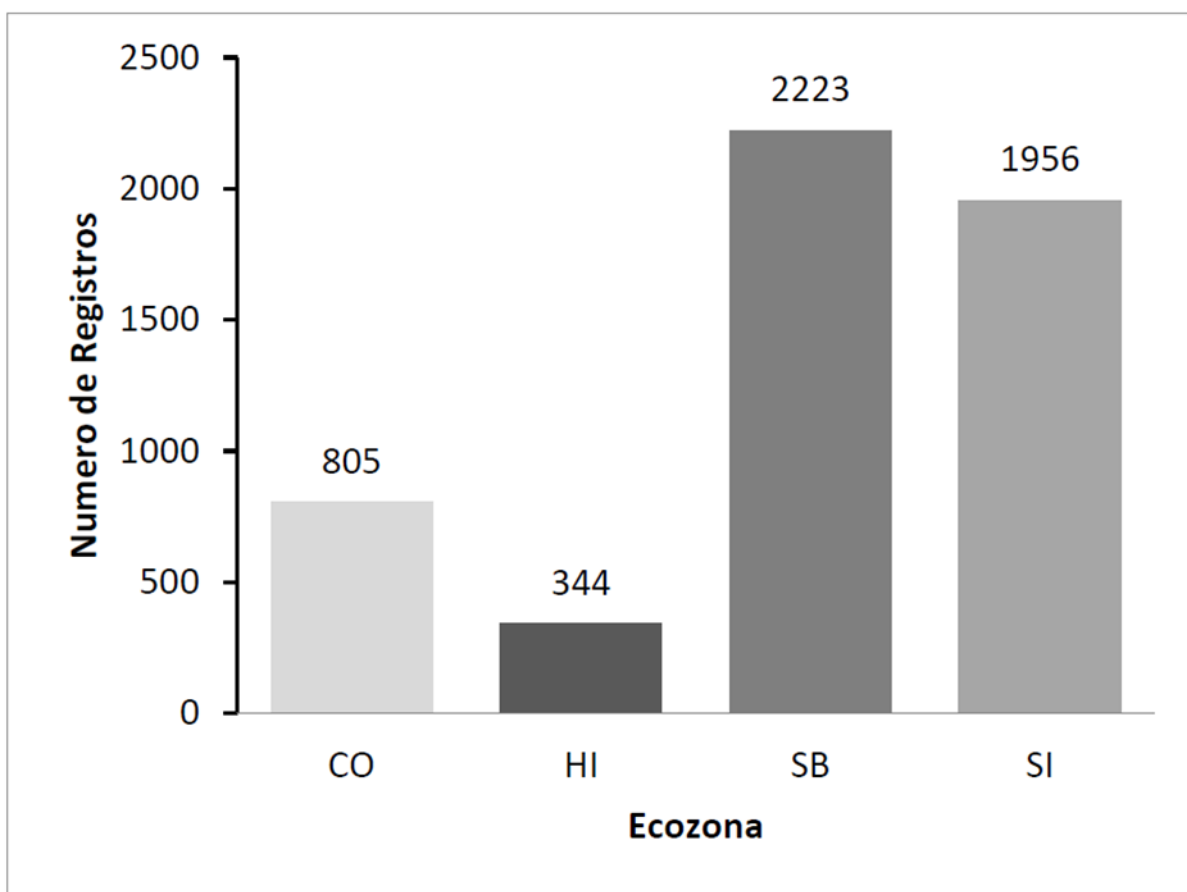
Keywords: Search engine, vector space model.

## **Capítulo 1**

### **EL PROBLEMA**

Actualmente la internet presenta una gigantesca ola de información, dicha herramienta es útil y necesaria para la formación, ya que posee múltiples servicios que permiten que la comunicación y difusión de la información sea efectiva a nivel mundial. Dicho fenómeno ha sido adoptado por las organizaciones que están comprometidas con la conservación de la biodiversidad, así mismo, se están desarrollando nuevos campos de investigación en la biodiversidad y biotecnología. Aunque queda mucho por explorar, organizaciones con plataformas en la web como GREENPEACE con presencia en 51 países, Global Biodiversity Information Facility con 1,703,252 especies, Banco de imágenes de la Facultad de Biología UCM, entre otros, han optado por hacer difusión usando herramientas tecnológicas.

Actualmente en nuestro país según SERFOR, (2016) en su informe parcial del inventario nacional forestal y de fauna silvestre, cuenta con una riqueza de 5328 registros de especies. De los cuales la ecozona Selva Baja fue la que tuvo mayor cantidad de registros con 2223, seguido por la ecozona Sierra con 1956 registros, luego la ecozona Costa con 805 registros y por último la ecozona Hidromórfica con 344 registros. Ciertas cifras de recursos naturales son de gran importancia científica, social, económica, ecológica y cultural. Es por eso que debe ser conservada como parte del patrimonio nacional, y este deber es responsabilidad de todos.



*Figura 1 Total de registros por zonas en el Perú*

Fuente: (SERFOR, 2016).

Asimismo, en el ámbito regional, en San Martín, se cuenta con instituciones que dedican parte de su trabajo al desarrollo de la investigación científica y tecnológica. El Instituto de Investigaciones de la Amazonía Peruana (IIAP) es una de ellas, ya que es su afán por concientizar y promover la conservación de la biodiversidad, coopera con las concesiones que tienen el mismo interés en particular, y uno de sus tantos esfuerzos es realizar inventarios de especies de flora y fauna en la Región. Los inventarios aparecen como una disposición del Servicio Nacional de Áreas Naturales Protegidas por el Estado – SERNANP, que es una institución pública unida al Ministerio del Ambiente desde mayo del 2008. Su cometido es garantizar la conservación de las Áreas Naturales Protegidas del país, su diversidad biológica y el mantenimiento de sus servicios ambientales. El requisito es impuesto a aquellas

concesiones de conservación, ya que el cumplimiento garantiza su sostenibilidad y consideración dentro de la categoría de conservación.

Sin embargo, según el Biol. Ángel Martín responsable del PIBA en el IIAP San Martín, en una entrevista nos comentó que el alcance no es el óptimo en este trabajo, ya que existen diversos factores relacionados con la difusión de la información científica, factores como la forma de difundir, los medios de comunicación que se utilizan, y la disponibilidad de recursos financieros, hacen que parte de este proceso sea deficiente.

El IIAP por su parte, realiza la difusión de los resultados de forma documental mediante folletos impresos, catálogos y libros, difundidos en eventos sociales (talleres de difusión), eventos de investigación (expo amazónica), y publicación en revistas científicas, entre otras. Por otro lado, no existe una difusión a iniciativa del interesado, pues la información no se encuentra almacenada y organizada digitalmente para su consulta ya sea presencial o en línea. Esta situación problemática afecta no solo a las instituciones que realizan investigación en biodiversidad, sino también a la comunidad científica en general.

Como alternativa de solución para esta problemática, surge la idea de implementar un motor de búsqueda basado en el modelo de espacio vectorial, lo cual dentro del ámbito internacional algunos investigadores adoptaron este método. Estos son tan solo algunos ejemplos de caso de éxito:

Según Martínez, (2002) en su tesis doctoral: “Propuesta y desarrollo de un modelo para la evaluación de la recuperación de información en internet”, Concluyó lo siguiente:

Los trabajos de evaluación de los SRI tienen mucha relación con la investigación y el desarrollo de la recuperación de información, es la propia base de estos sistemas la que motiva esta necesidad crítica de evaluación, del mismo modo como acontece en cualquier otro campo de trabajo que aspire a ser cualificado y certificado como científico.

Por otro lado, Barragán, (2008) en su trabajo de Investigación: “Un Método para

Recuperación de Información en Documentos Orales basado en Codificación Fonética”, concluye:

La codificación fonética es de vital importancia en la recuperación de información de documentos orales. Así mismo se logró aplicar dos métodos de codificación fonética Soundex y DM Soundex, los resultados de recuperación de información (RI) del segundo método fueron superiores con el primero, confirmando así que la codificación fonética al trabajar en conjunto con otras técnicas de RI mejoran los resultados de recuperación.

A su vez, Huacani, (2014) en su tesis: “Recuperación de la información empleando el modelo de espacio vectorial en la gestión documentaria para la unidad de resoluciones de la universidad nacional del altiplano – puno”, concluyo que:

El Modelo de Espacio Vectorial en la búsqueda de información es adecuada y oportuna y versátil cuando es para búsquedas con términos y consultas especializadas como el caso de las resoluciones, generando rankings de precisión fundamentalmente cuando se tiene grandes cantidades de registros en la base de datos. Además, también menciona que, el implementar la búsqueda de información, nos encamina a continuar haciendo aplicaciones de búsqueda más eficientes que permitan facilitar información a los usuarios de manera oportuna y precisa.

Por otro lado, Zambrano, (2009) en su trabajo de investigación WWW.BIOIMAGENES.COM cuyo objetivo principal fue diseñar y realizar un sitio Web dedicado a la fotografía documental ecológica, tomando como motivo inicial al mono Araguato. Menciona que, la conjugación de trabajos fotográficos y de investigación genera un importante contenido donde se pretende crear credibilidad dentro de los grupos a los que se dirige. Además, a raíz de esto proponer soluciones o simplemente aprender a cerca de realidades que vive la fauna y flora venezolana junto a su población humana.

Así mismo, afirma que, los ataques incesantes hacia el medio ambiente y su vida vegetal y animal no se reducen ni se aplacan con el pasar de los años; al contrario, con el paso del tiempo son más las zonas naturales afectadas, de que crecen los números de especies amenazadas por extinción, pero también, afortunadamente crecen los grupos que pretenden atender las necesidades de la madre tierra.

## **Capítulo 2**

### **REVISIÓN DE LA LITERATURA**

#### **2.1. Recuperación de información**

La humanidad ha carecido de medios sobre los cuales representar todo lo que acontece en el mundo y de difundir de alguna manera su progresivo crecimiento. El escribir sobre una hoja de papel, siempre ha sido el método “tradicional” y vital que tiene sostenibilidad del conocimiento en el tiempo. Este progreso ha permitido que la escritura sea representada de múltiples formas, teniendo una sostenida permanencia en el tiempo, donde la información se representa digitalmente y es posible almacenarla y distribuirla masivamente en forma sencilla y rápida, a través de los medios de comunicación. Los avances en la ciencia y tecnología ha permitido conocer nuevos horizontes en las que el hombre puede transformar la gran magnitud de datos en información valiosa, y la información en conocimiento (Bordignon & Tolosa Chacón, 2007).

La recuperación de la información (RI) considerado un campo de estudio bastante amplio y complejo cuyas definiciones están muy relacionadas con la digitalización de la información. Sin embargo, estudiar este campo fue posible gracias al aporte de Calvin Northrup Mooers en los años de 1950 quien presentó por primera vez el término. Y desde entonces la definición de RI ha ido evolucionando en los diferentes campos de la investigación.

Baeza & Ribeiro, (1999) consideran que, la RI es un método que permite ubicar un determinado conjunto de datos dentro de un repositorio de información o base de datos. Luego los resultados obtenidos, son el producto que ayuda al usuario a obtener la información requerida y esta es expresada normalmente en lenguaje natural.

Así mismo, en términos generales Bordignon & Tolosa Chacón, (2007) plantean que la RI intenta dar solución a la problemática de “encontrar y rankear documentos importantes que

complazca la necesidad de información de cada usuario, manifestado en un determinado lenguaje de consulta”.

## **2.2. Recuperación de la información en la world wide web**

### **2.2.1. Recuperación en la Web**

Según Merlino santesteban, (2003) la World Wide Web (WWW, W3 o Web) es una red hipertextual de múltiple complejidad que se incrementa a un ritmo acelerado. Esta enorme estructura provee patrones de conectividad que pueden mejorar progresivamente los procedimientos de búsqueda y ordenar por relevancia de los sistemas de búsqueda tradicionales.

Así mismo, Bordignon & Tolosa Chacón, (2007) manifiestan que la gigantesca ola de información tiene un crecimiento permanente y toma diferentes formas de representación, desde sencillos archivos de texto o periódicos electrónico hasta librerías digitales y espacios mucho más amplios y complejos como la web.

Sin embargo, la referencia web generalmente es más subjetiva y menos relevante que en la literatura científica. Los gestores de contenido web en muchas ocasiones no consideran la calidad, relevancia u objetividad de la información. Se podría decir que la creación de enlaces es un fenómeno esencialmente anárquico (Merlino santesteban, 2003).

### **2.2.2. Técnicas de recuperación en la Web**

Blázquez Ochando, (2013) lo define como el conjunto de procesos destinados a la recuperación de información, que abarca desde la producción de las colecciones, su depuración, indexado, procedimiento textual, clasificación, almacenamiento, recuperación a través de algoritmos booleanos, vectoriales, probabilísticos, todos estos fundamentados en el lenguaje, así como todos aquellos factores que incurren en cualquier aspecto relacionado como por ejemplo la interfaz de consulta, la conducta del usuario, la retroalimentación de las consultas y hasta la forma de cómo se presenta la información.



### **2.3. Modelos de recuperación de información**

Los modelos de recuperación se clasifican en Clásicos y Estructurales. En esta sección presentaremos los modelos clásicos.

#### **2.3.1. Modelo Booleano**

Este modelo está basado en la famosa teoría de conjuntos y el álgebra booleana, es el modelo de recuperación más simple. Su estrategia de recuperación consiste en una norma de decisión binaria (pertinente o no pertinente) sin ninguna noción de escala de medida, sin noción de una unión parcial en las condiciones de la consulta.

En el modelo booleano la representación de una colección de documentos es posible gracias a las matrices binarias, es decir “documento – término”, donde los términos han sido extraídos de forma manual o automática de los documentos y representan su propio contenido (Bordignon & Tolosa Chacón, 2007).

Del mismo modo, las consultas del usuario se construyen con términos relacionados por operadores lógicos (AND, OR, NOT) y las respuestas están relacionadas a documentos donde su representación cumple las condiciones lógicas de la expresión de búsqueda. En el modelo original no existe una lista indicando un nivel de importancia sobre el conjunto de respuestas a una consulta, al contrario todos los documentos poseen la misma importancia (Bordignon & Tolosa Chacón, 2007).

#### **2.3.2. Modelo Vectorial**

Este modelo de recuperación fue propuesto por primera vez por Gerard Salton dentro del marco del proyecto Smart. Este patrón hace posible que los documentos puedan ser vistos como vectores de términos, facilitando a los documentos ubicarse en un espacio vectorial de  $n$  dimensiones, o sea, con tantas dimensiones como elementos tenga el vector (Luque Rodríguez, 2005).

Dicho de otro modo, este modelo utiliza una matriz “Término – Documento” que comprende el conjunto de términos de la colección de referencia y los documentos existentes. En la coincidencia de un término  $t$  y un documento  $d$  se almacena un valor numérico de mucha importancia del término  $t$  en el documento  $d$ ; dicho valor representa su poder de discriminación. De este modo, cada documento puede ser visto como un vector que pertenece a un espacio  $n$ -dimensional, donde  $n$  es el número de términos que hacen realidad el vocabulario de la colección. En teoría, los documentos en cuyo contenido se logre identificar términos con un alto nivel de similitud estarían a muy poca distancia entre sí sobre tal espacio. Del mismo modo se procede con la consulta, que es un documento adicional y se realiza un mapeo sobre el espacio de documentos. Después, teniendo la consulta formulada es posible devolver una lista ordenada de documentos por distancia (los más importantes primero). Para realizar el cálculo de similitud entre el vector consulta y los vectores que representan los documentos se emplean múltiples fórmulas de distancia, llegando a ser la más conocida la del coseno (Bordignon & Tolosa Chacón, 2007).

### **2.3.3. Modelo Probabilístico**

Este modelo se explica a través de las probabilidades. Esto dependerá de la frecuencia de aparición de un término en un documento o conjunto de documentos podría considerarse un dato relevante al momento de hacer una consulta. A continuación, presentaremos un ejemplo:

En la siguiente figura de consulta se puede separar una colección de  $N$  documentos (figura 2) en cuatro grupos diferentes: REL conjunto de documentos relevantes, REC

conjunto de documentos recuperados, RR conjunto de documentos relevantes recuperados y NN el conjunto de documentos no relevantes no recuperados (Bordignon & Tolosa Chacón, 2007).

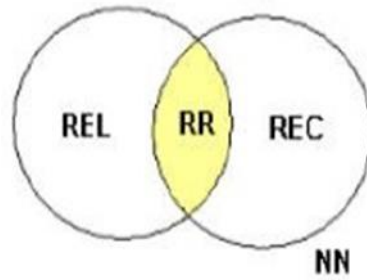


Figura 2 Modelo de relevancia probabilístico

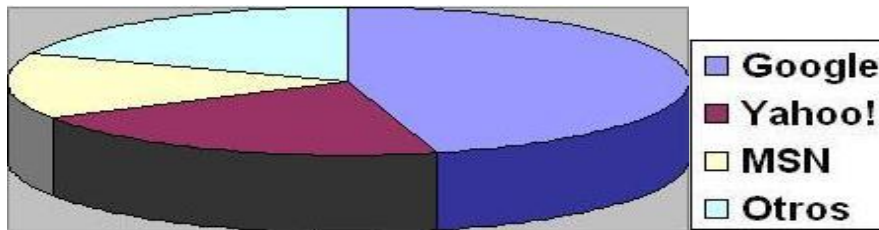
#### 2.4. Modelo de Espacio Vectorial (MEV)

Para Martínez, (2006) el modelo de espacio vectorial consiste en calcular el grado de similitud de una consulta realizada por un usuario con respecto a los documentos de la colección cuyos términos fueron ponderados mediante TF-IDF. El Modelo fue presentado por Salton en 1975 y posteriormente perfeccionado en 1983 junto con Mc Gill, basándose en tres principios esenciales:

- La proporción parcial, esto es, la suficiencia del sistema para organizar los resultados de una búsqueda y sobre todo ordenarlos, basado en el grado de similaridad entre cada documento de la colección y la consulta.
- La ponderación de los términos en los documentos, no limitándose a señalar la presencia o ausencia de los mismos, sino atribuyendo a cada término y en cada documento un número real que muestre su relevancia en el documento.
- La ponderación de los términos en la consulta, de manera que el usuario puede asignar pesos a los términos de la consulta que muestren la relevancia de los mismos en relación a su necesidad informativa.

A si mismo menciona que, en el MEV los documentos y las consultas se muestran mediante un conjunto de números ordenados, y no solamente de ceros y unos como se menciona en los modelos anteriores; sin embargo, este modelo considera al cero como la ausencia del término en el documento.

Existen algunos datos determinantes sobre los buscadores Web afirman que, entre el 80% y el 95% de las búsquedas se realizan con Google, Yahoo! y MSN.



*Figura 3 Buscadores más populares de la historia*

El modelo vectorial en sus inicios fue utilizado por Yahoo! como algoritmo de recuperación, ya que posee versatilidad y eficiencia al momento de generar rankings de precisión en colecciones inmensas, esto hace especial al modelo para determinar la equivalencia parcial de los documentos.

Por otro lado, es importante considerar que la ponderación de los términos se efectúa mediante los factores TF y IDF.

#### **2.4.1. Factor Term Frequency (TF)**

La frecuencia de aparición de un término, es el resultado de sumar todas las ocurrencias o la cantidad de veces que aparece un término en un documento. A esta clase de frecuencia de aparición también se la nombra “Frecuencia de aparición relativa” por que atañe a un documento en específico y no a toda la colección.

Tabla 1  
*Características del Factor TF*  
*Fuente: (Blázquez Ochando, 2012)*

Factor TF	
Denominación	Term Frequency = Frecuencia de aparición del término
Descripción	Es la frecuencia de veces que un término aparece en todo un documento, Esto permite poder determinar su capacidad de representatividad.
Finalidad	Representativa. Frecuencia de aparición TF baja.
Casos	Representatividad elevada. Frecuencia de aparición TF media.

El cálculo se realiza una vez que el texto del documento haya ingresado al proceso de normalización, respondiendo a los principales procesos de discriminación mencionados. Como siguiente paso se realiza el conteo del número de veces que el término aparece presente en el documento. Así mismo, se deja bien claro que es vital realizar el cálculo de TF de cada término en cada documento. Actualmente utiliza un algoritmo de mejorado denominado WebRank.

$$tf(n) = \sum_{D1} (n)$$

La frecuencia de aparición de un término (n) en un documento (D1) es la suma de las ocurrencias de dicho término

Figura 4 Cálculo de la frecuencia de aparición

### 2.4.2. Factor Inverse Document Frequency (IDF)

La frecuencia Inversa del Documento para un término, es inversamente proporcional al número de documentos en los que aparece dicho término. Esto indica que cuanto menos sea la cantidad de documentos, así como la frecuencia absoluta de aparición del término, mayor será su factor IDF y a la inversa, cuanto mayor sea la frecuencia absoluta relativa a una alta presencia en todos los documentos de la colección, menor será su factor de discriminación.

Tabla 2

*Características del factor IDF*

*Fuente: (Blázquez Ochando, 2012)*

---

Factor TF	
Denominación	Inverse Document Frequency = Frecuencia Inversa del Documento para un término
Descripción	Es el coeficiente que determina la capacidad discriminadora del término de un documento con respecto a la colección. Es decir, distinguir la homogeneidad o heterogeneidad del documento a través de sus términos.
Finalidad	Discriminatoria.  Poder discriminatorio bajo. El término es genérico y aparece en la mayoría de los docs.
Casos	Poder discriminatorio medio.,  Poder discriminatorio alto. El término es especializado y aparece en pocos docs.

---

El factor IDF es exclusivo para cada término de la colección. Es decir, que su cálculo, véase figura2, el IDF de un término dado (n) se consigue aplicando el logaritmo en base 10 de N (Número total de documentos de la colección) dividido entre la "Frecuencia de documentos para un término (n) en la colección" (o lo que es lo mismo el número de documentos de la colección en los que aparece el término (n) dado). Al resultado se le suma 1

para enmendar los valores para los términos con IDF muy bajos (Estos cambios están sujetos al sistema de recuperación).

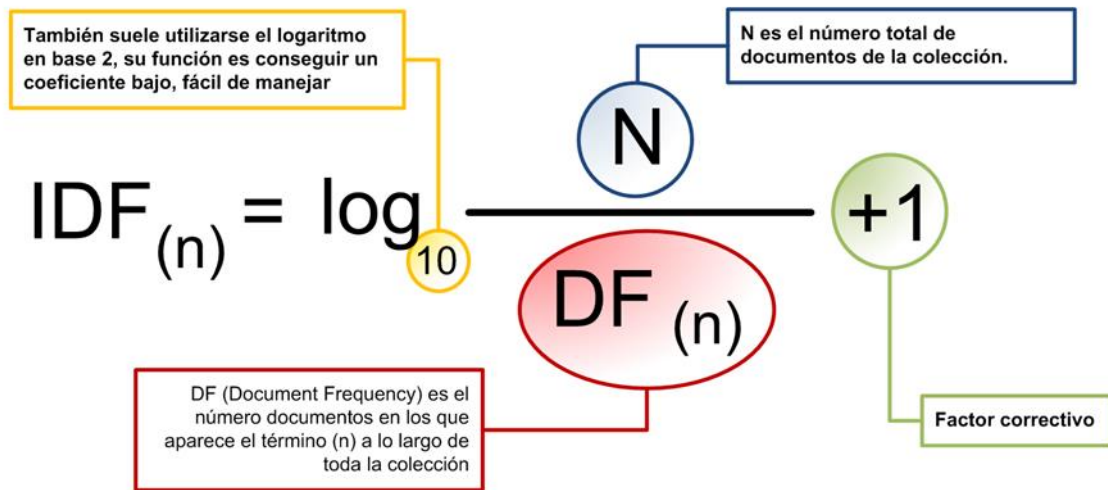


Figura 5 Cálculo del IDF de un término

### 2.4.3. Ponderación TF-IDF

El peso de un término en un documento es el producto de su frecuencia de aparición en dicho documento (TF) y su frecuencia inversa de documento (IDF)

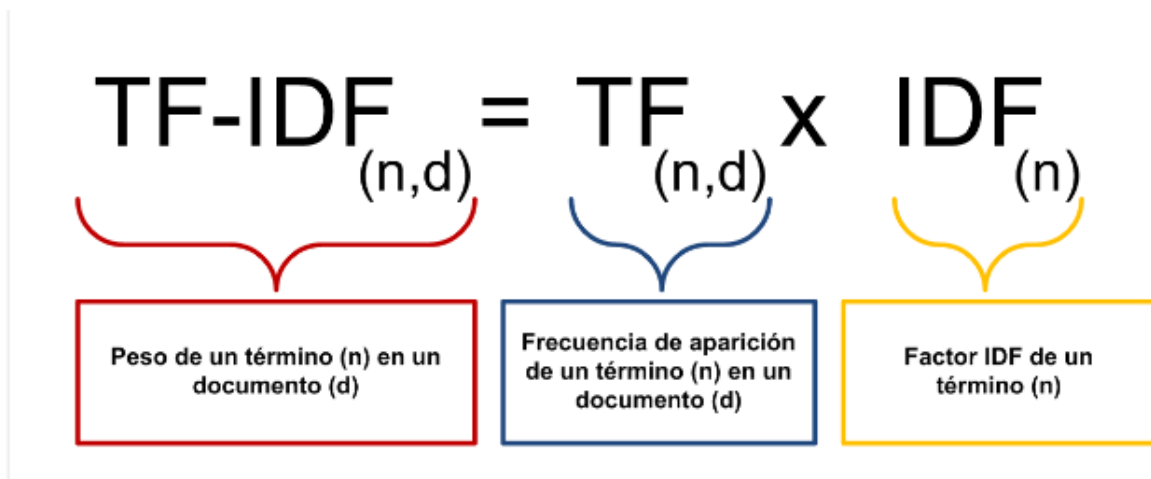


Figura 6 Ponderación TF-IDF

## 2.5. Motor de búsqueda

Comúnmente son llamados buscadores, motores de navegación o motores de búsqueda, conjunto de programas o herramientas interactivas que agilizan la búsqueda y recuperación de información en Internet. Los motores de búsqueda (MB) muestran formularios donde cualquiera puede ingresar los datos a través de una interfaz de mucha sencillez y fácil comprensión para el usuario, el cual escribe una palabra clave o frase y obtiene una lista de recursos que se corresponden con el criterio indicado. Los MB no pueden abarcar todos los recursos disponibles en Internet, sin embargo, muchos contienen referencias a millones de recursos. Los resultados, por tanto, pueden variar de un motor de búsqueda a otro (Rodríguez Camiño, 2003).

En síntesis, un MB es una plataforma virtual que permite recuperar datos o archivos almacenados en un servidor de Internet.

A continuación, presentamos una infografía de algunos buscadores web que son tendencia mundial en cuanto a su uso, durante diciembre del año 2017 al diciembre del 2018.

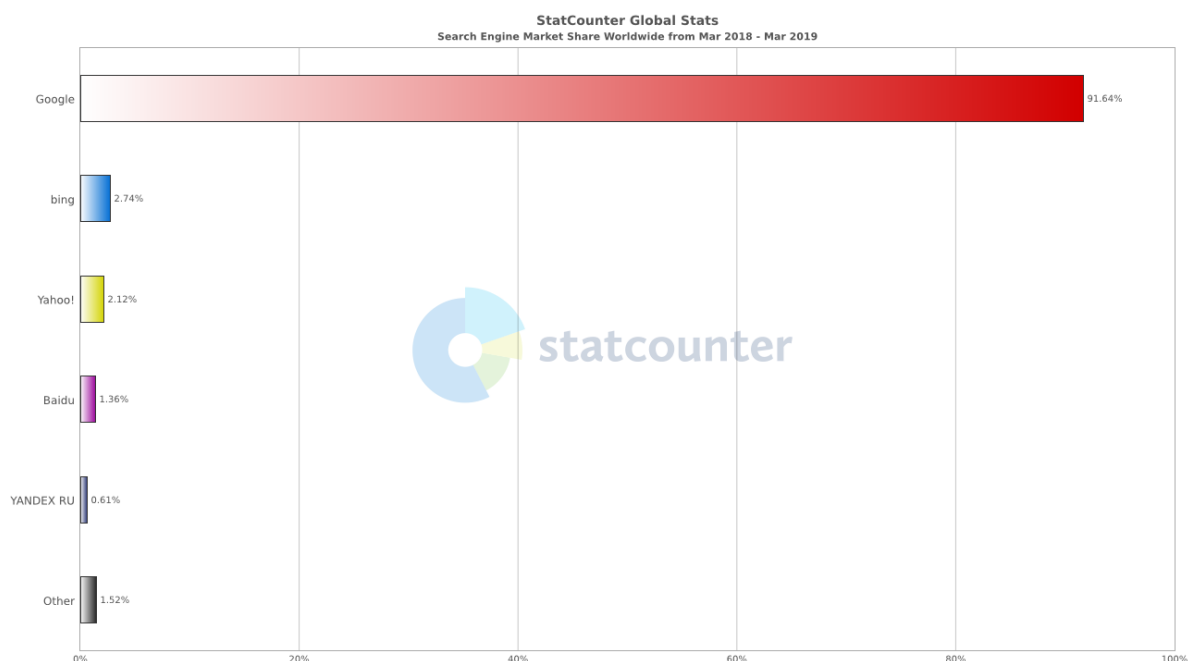


Figura 7 Search Engine Market Share Worldwide Mar 2017 – Mar 2018

Fuente:(StatCounter Global Stats, 2019)



## **2.6. Difusión de la información**

Es un proceso que permite transmitir al usuario la información que necesita o darle la posibilidad de obtenerla. Así mismo, es importante resaltar que la información que se genera y recolecta cada día tiene que ser difundida cuidadosamente, para el desarrollo del conocimiento (Castillo Blasco, 2014).

### **2.6.1. Formas de difusión**

No existe una sola forma de difusión, existen múltiples maneras e inclusive servicios con la capacidad de canalizar la información hacia los usuarios.

Referente a los medios de difusión algunos son: papel impreso, tabloneros de anuncios, soporte magnético para consulta en ordenadores, páginas web, correo electrónico, difusión verbal (persona a persona, cursos), medios audiovisuales como videos informativos (Castillo Blasco, 2014)

Existen dos formas básicas de difundir la información como, la difusión bajo demanda y la documental.

### **2.6.2. Difusión bajo demanda**

En este caso el usuario toma la iniciativa y se dirige al centro de datos y solicita la información. El documentalista consulta al sistema y obtiene la información de interés. Las búsquedas bibliográficas y documentales logran contestar a esta necesidad de información propuesta por el usuario. Las búsquedas pueden ejecutarse en bases de datos documentales elaboradas por el propio centro, pero también, sobre bases externas.

Se define también a la difusión bajo demanda como la consulta en línea a una Base de datos por parte del usuario sin la mediación de un documentalista (Castillo Blasco, 2014).

### **2.6.3. Difusión documental**

En este caso el centro de datos es el que oferta contenido al usuario de acuerdo a un estudio de sus necesidades como, que tipo de información necesitan, y como lo necesitan, así

mismo, es objeto de estudio el sistema de difusión, para determinar si es adecuado para cada necesidad (Castillo Blasco, 2014).

La difusión documental además de ser una difusión activa, se clasifica de la siguiente manera:

### **2.6.3.1. *Difusión de documentos primarios***

#### **2.6.3.1.1. *Difusión directa en el centro***

Para el caso del acceso controlado. El usuario tiene permitido realizar consultas a los ficheros, pero debe solicitar el documento que necesite a un empleado del centro. Este principio se emplea para las colecciones que cuentan con documentos especiales por su contenido o soporte, obras raras o únicas, materiales débiles, materiales valiosos.

Cuando el acceso es libre. Los documentos siempre están disponibles para el usuario, y se encuentran ordenados generalmente según criterios temáticos. Este método propicia más libertad para el usuario, pero disminuye el control de los fondos y requiere más espacio.

#### **2.6.3.1.2. *Prestamos***

El usuario tiene la oportunidad de obtener el documento por un periodo de tiempo, para luego devolverlo, por otro lado, existe la posibilidad de obtener una copia si es que así lo desea el usuario.

### **2.6.3.2. *Difusión de referencias de documentos***

Un gran número de centros de documentación elabora algún tipo de publicación secundaria con muchas referencias de documentos y busca transmitir a los usuarios para su conocimiento.

#### **2.6.3.3. *Difusión selectiva de la información***

Conocido comúnmente como la difusión “a la carta”, por el cual de forma exclusiva se oferta a cada usuario las referencias de documentos que se relacionan con sus temas de

interés seleccionados a partir de todos los documentos recibidos durante un determinado período.

## **2.7. Biodiversidad**

La humanidad vive en un escenario en el cual las amenazas a la biodiversidad en sus diferentes niveles se vienen incrementando de manera más conspicua, por ejemplo la Unión Mundial para la Naturaleza (IUCN) calcula que en la actualidad hay 7000 especies de animales en situación de peligro, mientras que el número de las vegetales en la misma situación se acerca a las 80003; así también en La Evaluación de los Ecosistemas del Milenio los expertos confirman la degradación de 15 funciones de ecosistemas de un total de 24 (Romero, 2005).

Se define a la Biodiversidad como la variedad de seres vivos que habitan en la tierra y los patrones naturales que lo conforman. Así mismo, abarca la gama de ecosistemas, de especies y de sus poblaciones, así como las diferentes genéticas entre los individuos que la constituyen (Jiménez Sierra, Torres Orozco, & Corcuera Martínez del Río, 2010).

La Biodiversidad contiene, una gigantesca variedad de formas, y a través de cada una de ellas se organiza la vida. Abarca todas y cada una de las especies que cohabitan con nosotros en el planeta, sean animales, plantas, virus o bacterias, los espacios o ecosistemas a las cuales pertenecen y los genes que hacen a cada especie, y dentro de ellas a cada individuo, diferente del todos los demás (Dorado Nájera, 2010).

## **2.8. Minería de datos**

Conformado por un conjunto de técnicas y tecnologías que permiten la exploración de las bases de datos, de la manera más automática, con el objetivo de encontrar patrones o reglas que muestran cómo se comportan los datos en un determinado escenario. Para ello existen diversos tareas y métodos que permiten resolver dichas tareas.

Las tareas se pueden definir como un tipo de problema a ser resuelta por un algoritmo de minería de datos. Existen dos tipos de tareas las predictivas cuyo objetivo es estimar valores futuros, y las descriptivas que permiten identificar patrones en los datos que los explican o resumen. Por otro lado, los métodos sirven para resolver más de una tarea, a través de algoritmos o métodos. Algunos métodos que permiten resolver son los siguientes: Técnicas algebraicas y estadísticas, técnicas algebraicas, técnicas basadas en conteos de frecuencia y tablas de contingencia, entre otros (Galán Cortina, 2015).

## **2.9. Modelo de procesos para proyectos de minería de datos**

Existe diversidad en los modelos de proceso propuestos para el desarrollo de proyectos de Minería de Datos tales como SEMMA (Sample, Explore, Modify, Model, Assess), DMAMC (Definir, Medir, Analizar, Mejorar, Controlar), y CRISP-DM (Cross Industry Standard Process for Data Mining) sin embargo, el CRISPDM es un modelo muy utilizado en los sector académico e industrial.

### **2.9.1. Modelo CRISP-DM**

Es una guía de referencia que facilita el desarrollo de proyectos de Minería de Datos. Tienen como fundadores al grupo de empresas SPSS, NCR y Daimler Chrysler en el año 2000. En la actualidad es muy útil en el desarrollo de proyectos de Data Mining. El Modelo organiza el proceso en seis etapas: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación y Despliegue. La secuencia de las etapas, no es necesariamente inflexible. Cada etapa es desglosada en varias tareas generales de segundo nivel. Las tareas genéricas se detallan a tareas más específicas, pero en ningún momento se menciona como desarrollarlas. Es decir, CRISP-DM define un conjunto de tareas y actividades para cada etapa del proyecto, pero no detalla cómo realizarlas (Moine, Haedo, & Gordillo, 2011).

La implementación de tecnologías en los negocios es un gran desafío ya que requieren de una metodología. La selección de una de ellas dependerá del proyecto a trabajar, Así también tendrá como punto de partida las experiencias, inclusive teniendo en cuenta al que tenga el mejor de los procedimientos exitosos y populares.

CRISP-DM es la metodología que lidera el mercado, esto según la encuesta que KDnuggets realizó en el 2014 en comparación con la del 2007, los resultados son sorprendentemente estables (KDnuggets, 2014).

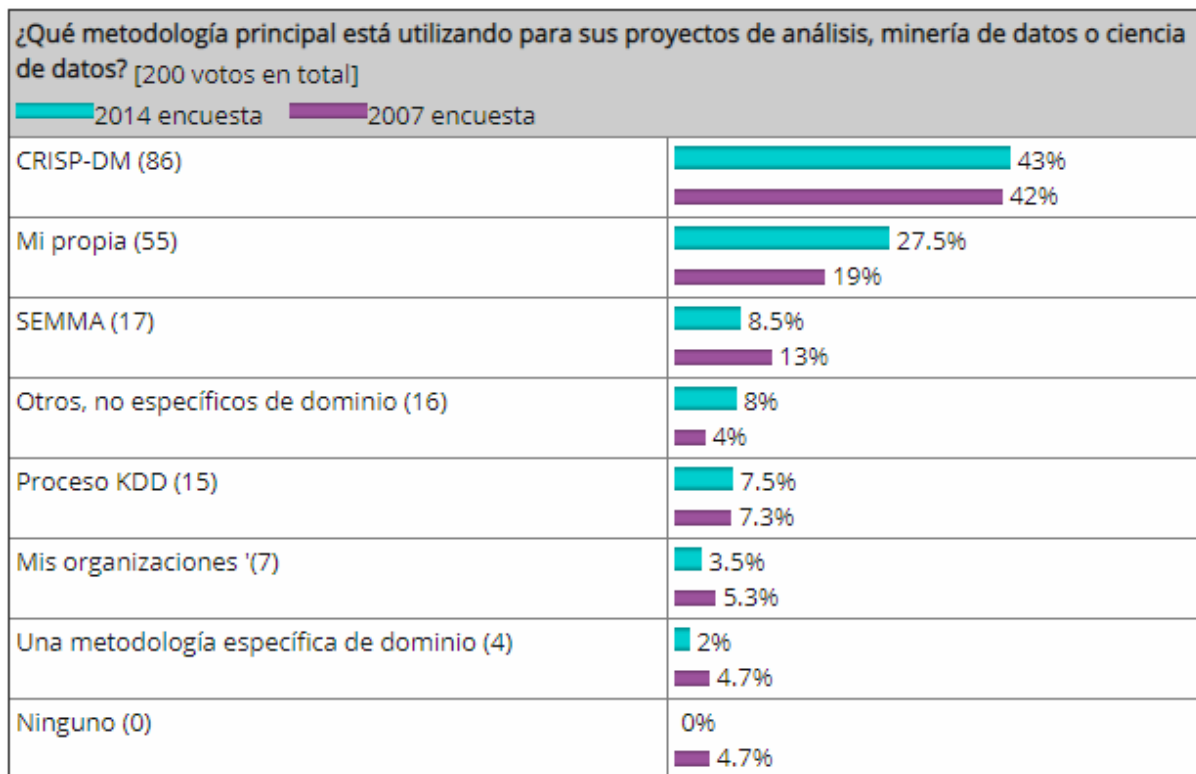
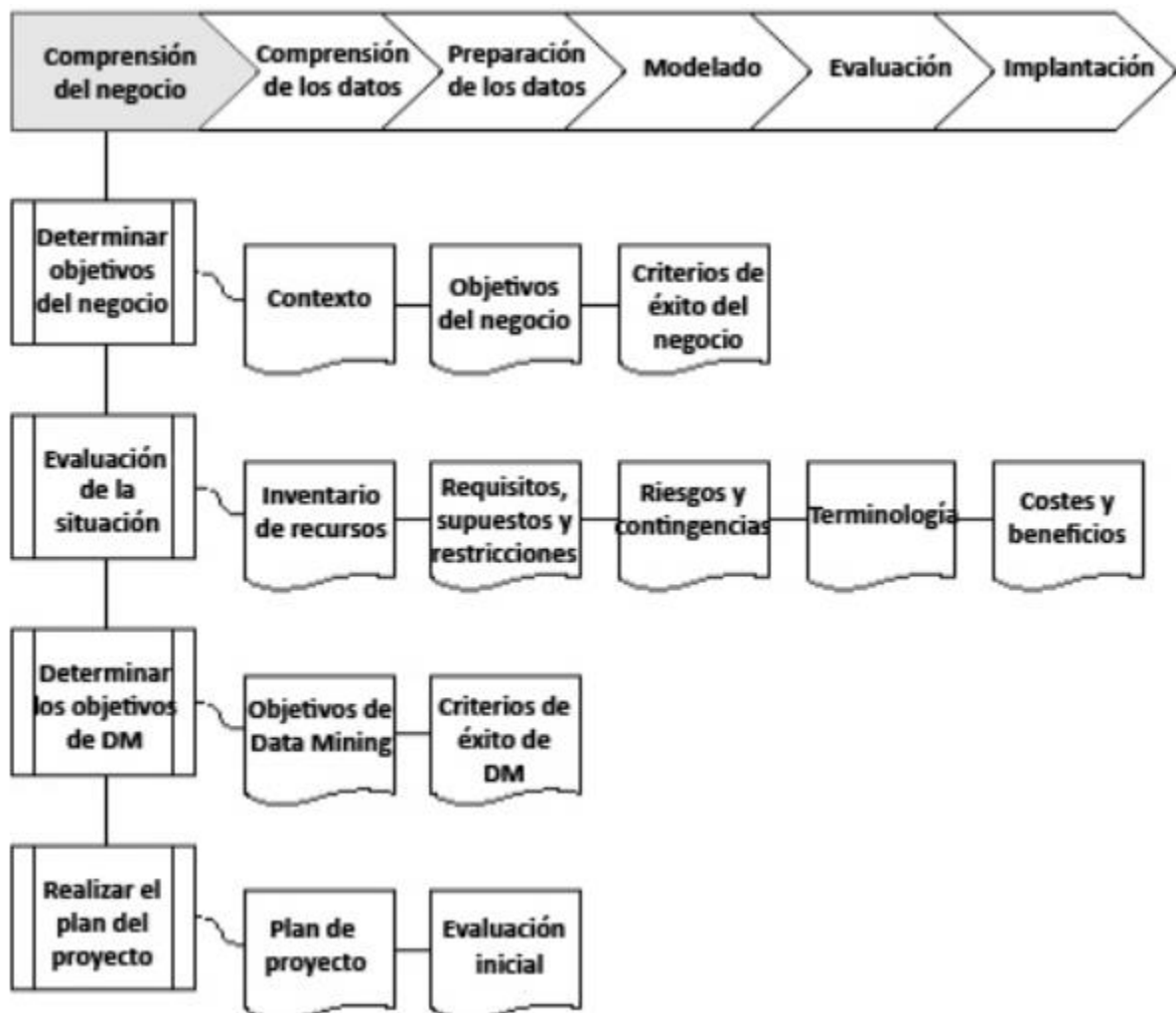


Figura 8 Metodologías más utilizadas del Data Mining  
Fuente: (KDnuggets, 2014)

A continuación, presentamos las fases de la gestión del proyecto, según (Chapman et al., 2000).

### 2.9.1.1. *Comprensión del Negocio*

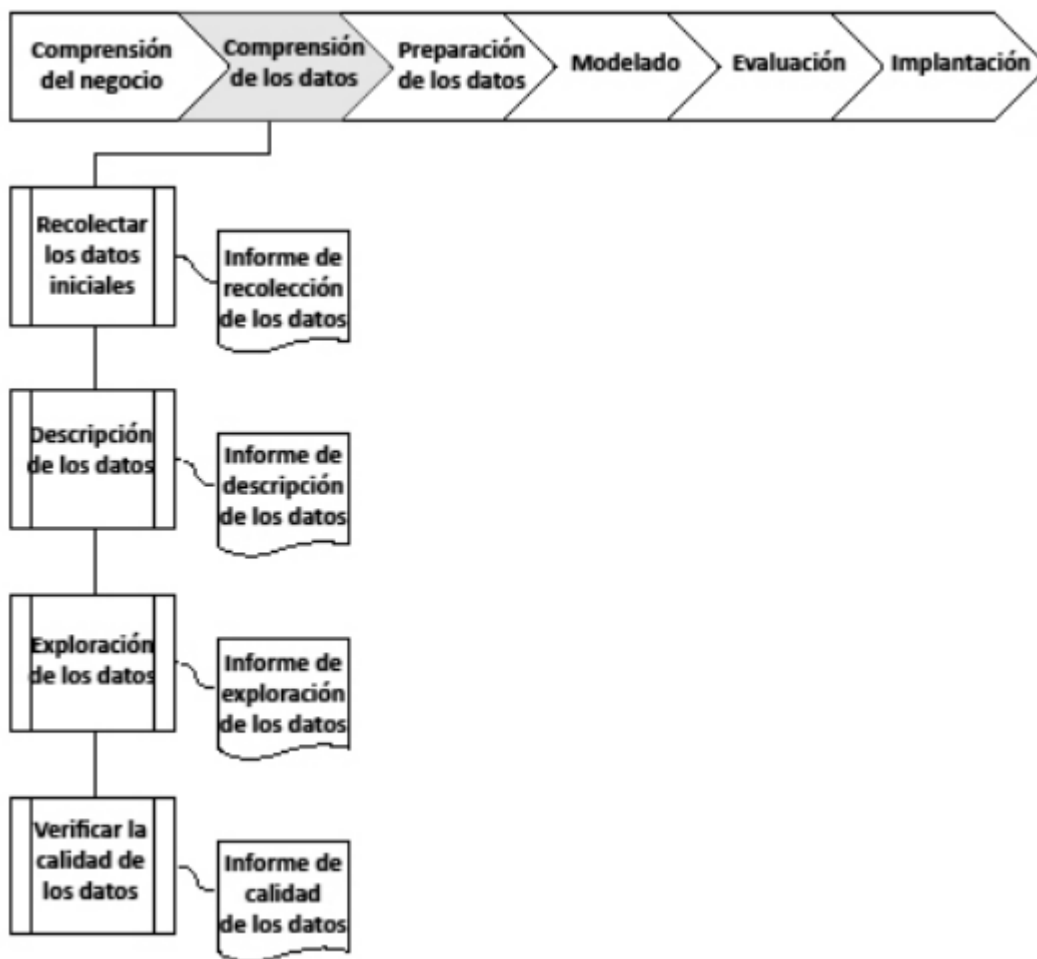
Es la más importante debido a que acopia tareas de comprensión de datos y requisitos del proyecto, a fin de transformarlos en objetivos técnicos y en un plan de proyecto. En esta etapa es de vital importancia convertir el conocimiento adquirido de negocio en un problema de minería de datos y luego en un plan preliminar cuya meta sea alcanzar los objetivos del negocio. A continuación, en la siguiente *Figura 9* se presentan las tareas de esta fase.



*Figura 9 Fase de comprensión del negocio*

### 2.9.1.2. *Comprensión de los datos.*

Esta fase comprende la recolección de datos iniciales con el objetivo de establecer un primer contacto con el problema y familiarizarse con ellos. La creación de una base de datos a partir de la corporativa es un hecho, ya que es posible que se generen frecuentes y abundantes accesos a la Base de Datos con el objetivo de hacer consultas. Las tareas comprenden esta fase son las siguientes, *Figura 10.*



*Figura 10 Fase de comprensión de los datos.*

### 2.9.1.3. Preparación de datos

En esta fase se procede a adaptar los datos a las técnicas de minería de datos que se van a utilizar posteriormente. Las tareas generales son las de proponer técnicas de modelado, limpieza de datos y cambios de formato. En la *Figura 11* se presenta a detalle las tareas

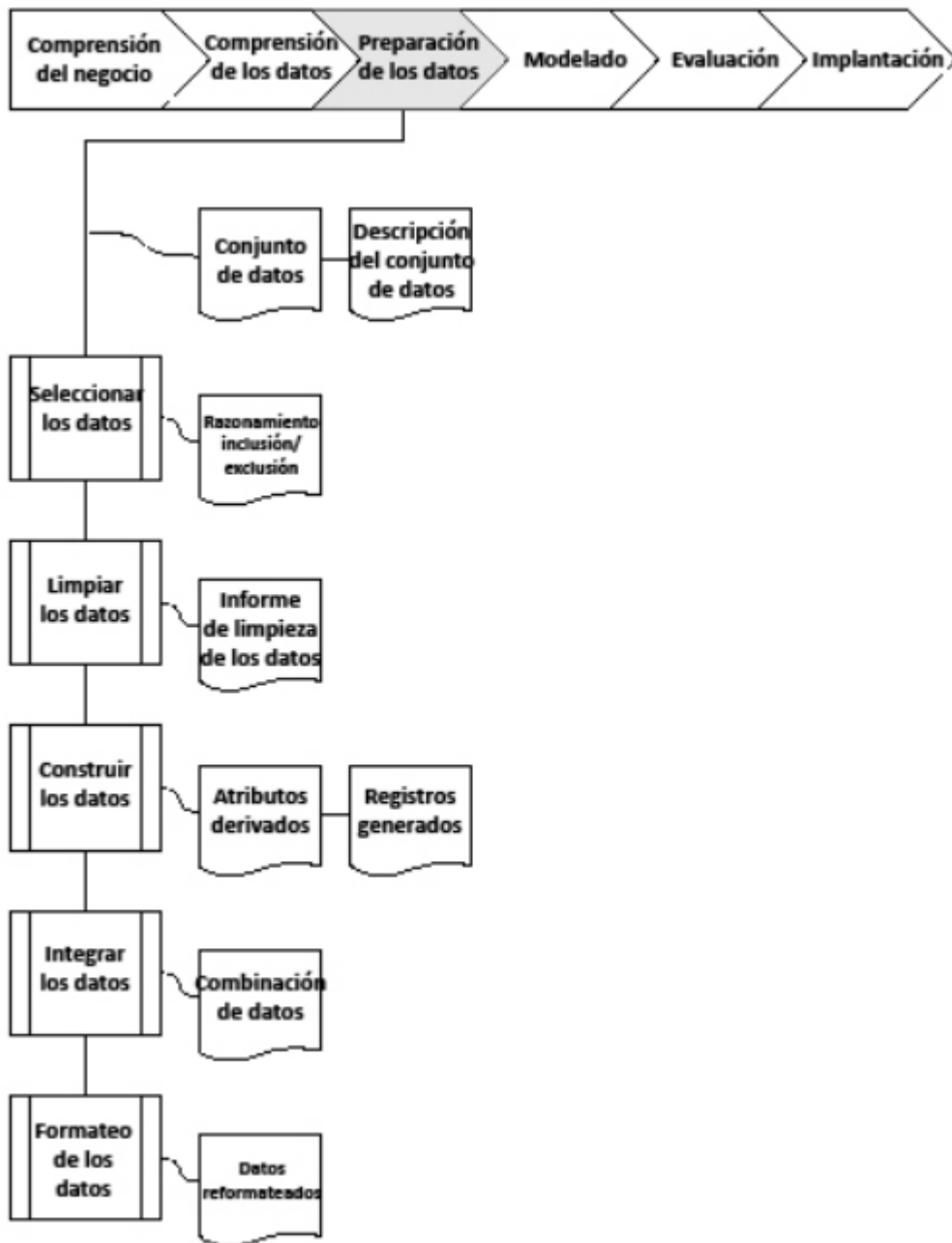


Figura 11 Fase de preparación de los datos



#### 2.9.1.4. Modelado

En quinta fase de se seleccionan las técnicas de modelado más adecuadas para el proyecto. La selección de las técnicas se basa en ciertos criterios definidos por la metodología. Después de estas tareas se genera y evalúa el modelo. La *Figura 12* muestran las tareas y las salidas que se obtienen en esta fase.

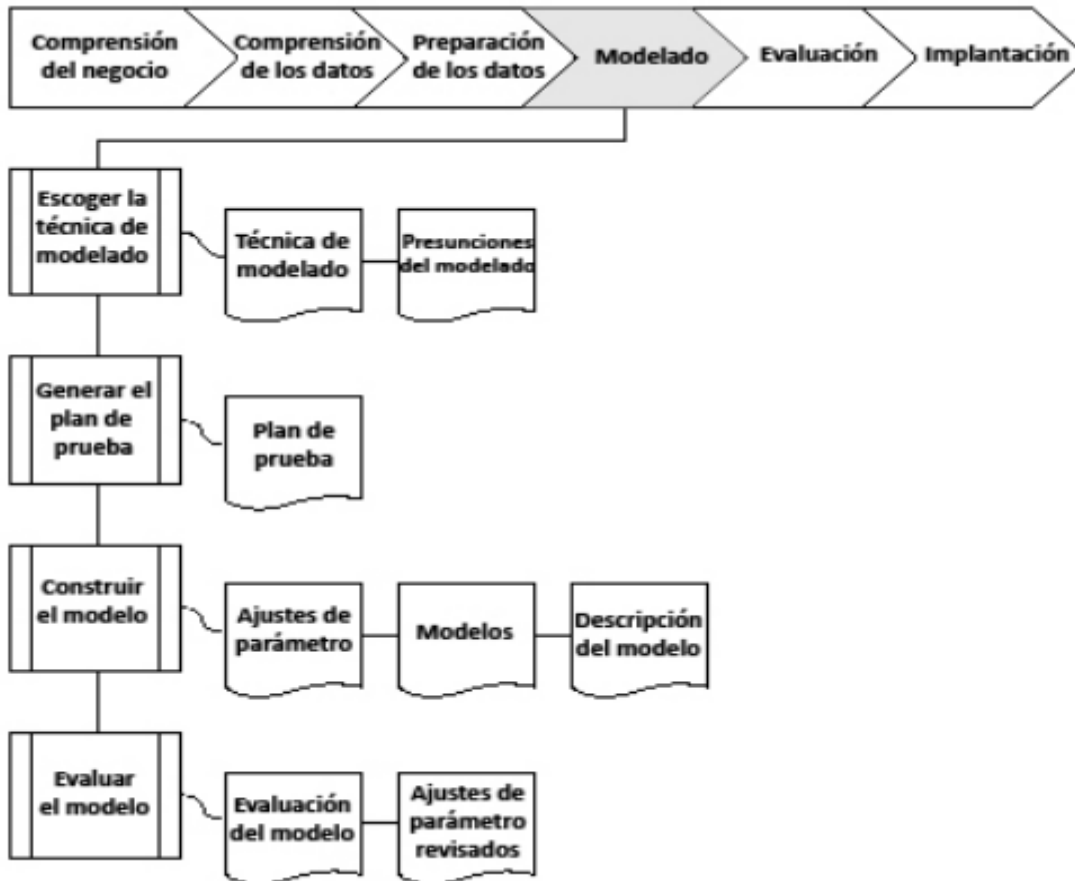
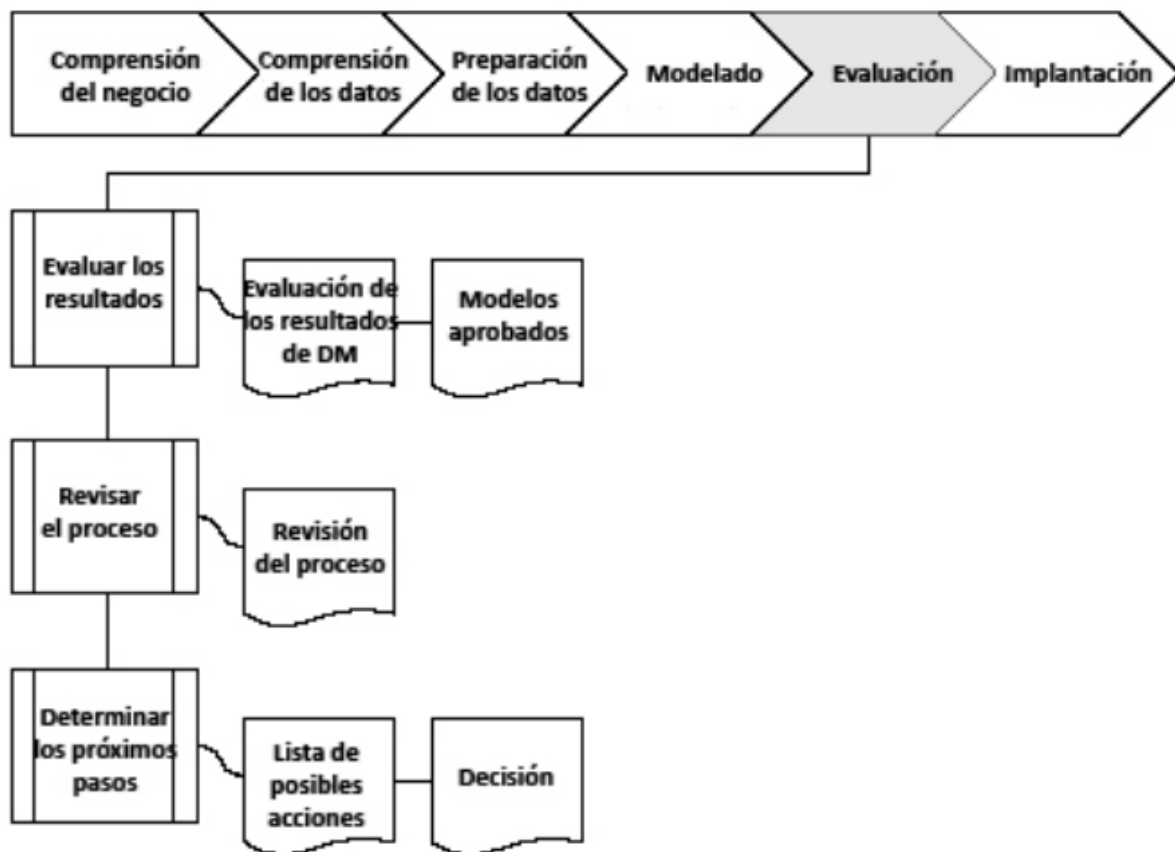


Figura 12 Fase de modelado

### 2.9.1.5. Evaluación

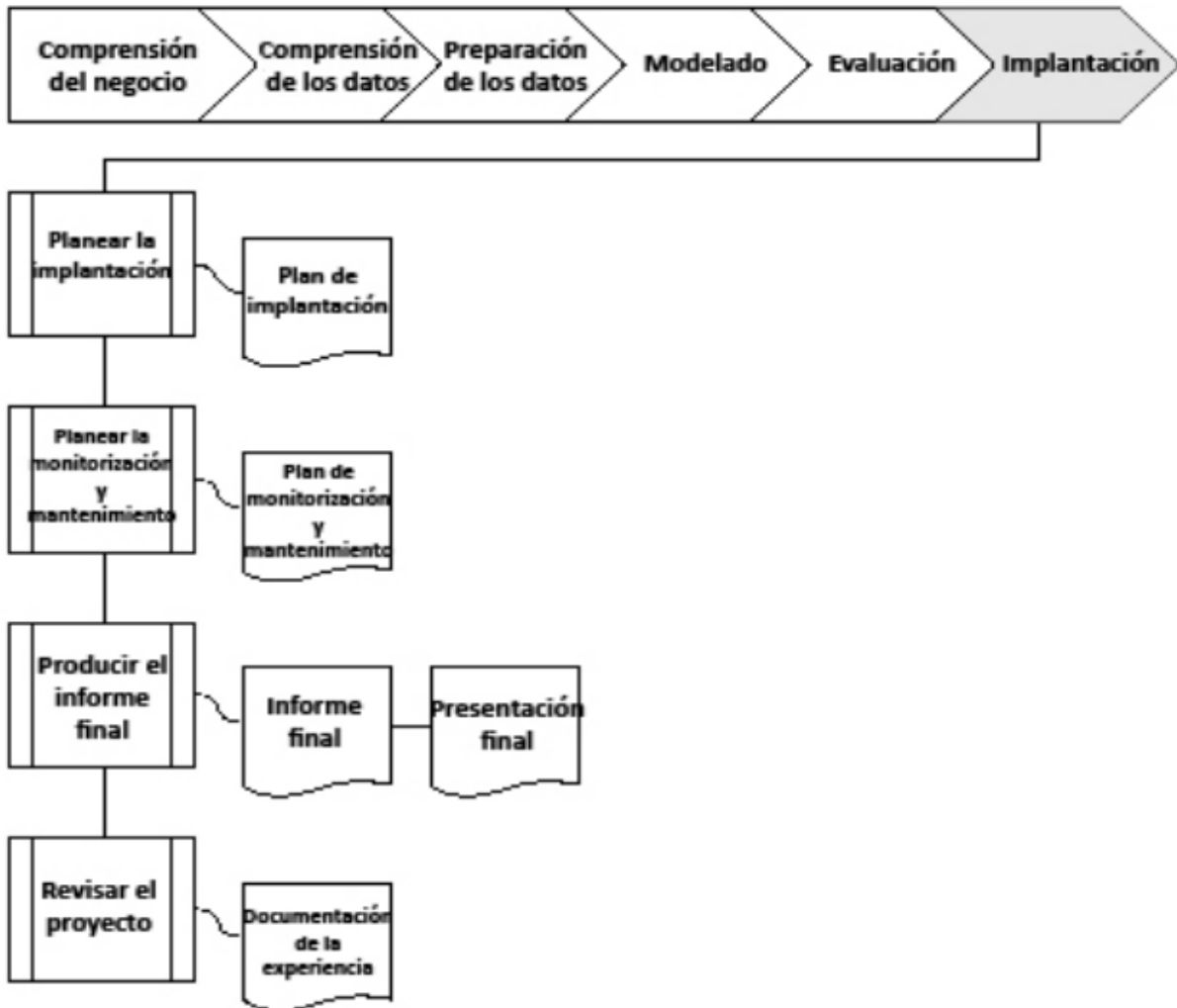
La evaluación del modelo tiene lugar en esta fase, considerando el cumplimiento de los criterios de éxito del problema. La revisión del proceso a partir de los resultados es indispensable, para garantizar que no ocurra incidencias en el futuro. El detalle se aprecia en la *Figura 13*.



*Figura 13 Fase de evaluación.*

### 2.9.1.6. Despliegue

Con el modelo elaborado y validado, el conocimiento conseguido se transforma en hechos dentro del negocio. Una recomendación de un analista basado en la observación del modelo y su resultado es un claro ejemplo. Los detalles de las tareas, están en la siguiente *Figura 14*.



*Figura 14 Fase de implantación*

En este sentido, los estudios confirman que la metodología para proyectos de Minería de Datos denominada CRISP-DM, es la más utilizada (Moine et al., 2011). Por lo tanto, consideramos que su sencillez, rendimiento y sostenibilidad en el mercado hacen que sea el más adecuado para abordar proyectos o extraer ideas para diseñar o revisar métodos de trabajo de similares características.

## **Capítulo 3**

### **MATERIALES Y MÉTODOS**

#### **3.1. Tipo de investigación**

Según Tacillo, (2016) sostiene que, la investigación aplicada consiste en emplear teorías y leyes científicas probadas, para explicar y dar solución a problemas de la realidad.

Este tipo de investigación permite solucionar problemas prácticos en los diferentes campos como la medicina, química, biología u otras disciplinas de la ciencia. El objetivo de la investigación aplicada es explicar teorías o predecir un comportamiento específico en una situación definida.

Es por eso que la investigación empírica a desarrollar es del tipo aplicada, ya que se buscamos aplicar el conocimiento adquirido con el fin de consolidar el saber para resolver el problema de difusión de la información.

#### **3.2. Descripción del lugar de ejecución**

La presente investigación se desarrollará en el Instituto de investigaciones de la Amazonía peruana (IIAP), ubicado en el Jr. Belén Torres de Tello N° 135 – Morales – Tarapoto – San Martín. El IIAP, es una institución de investigación científica y tecnológica, fundada para alcanzar el desarrollo sostenible de la población amazónica, teniendo en cuenta el aspecto rural, conservación y uso idóneo de los recursos naturales en la región amazónica. Está estructurada bajo el enfoque de seis programas de investigación: Biodiversidad amazónica (PIBA), Manejo integral del bosque y servicios ambientales (PROBOSQUES), Uso y conservación del agua y sus recursos (AQUAREC), Cambio climático, desarrollo territorial y ambiental (PROTERRA), Diversidad cultural y economía amazónica (SOCIODIVERSIDAD), por último, el programa en la cual estamos basando nuestra investigación, Información de la biodiversidad amazónica (BIOINFO). Todos estos aportan al manejo, uso sostenible, y conservación de la biodiversidad, y así también como al conocimiento de la sociodiversidad y economía amazónica.

Además, como institución cuenta con un campus principal en Loreto y con cinco campus ubicadas en las regiones amazónicas del Perú, Ucayali, San Martín, Madre de Dios, Amazonas y Huánuco, además cuenta con una oficina de coordinación en la capital del Perú.

### **3.3. Población y muestra**

#### **3.3.1. Población**

La población de estudio serán las instituciones que realizan investigación, recolectando y procesando información científica sobre la biodiversidad de especies en la región San Martín, con el objetivo de difundir y dar a conocer a la sociedad. Dichas instituciones son las siguientes:

- Administración Local del Agua.
- URKU Estudios Amazónicos.
- Proyecto Especial Huallaga Central y Bajo Mayo.
- Proyecto Especial Alto Mayo.
- Instituto Nacional de Innovación Agraria.
- Cima Cordillera Azul.
- EMAPA San Martín.
- Servicio Nacional de Áreas Naturales Protegidas por el Estado.
- Unidad de Gestión Educativa Local.
- Instituto de Investigaciones de la Amazonía Peruana.
- Dirección Regional de Agricultura San Martín.
- Servicio Nacional de Sanidad Agraria.

#### **3.3.2. Muestra**

Para Hernández et al., (2014) la muestra es la segmentación de la población o universo. Y con mucha precisión tienen que definirse y delimitarse de antemano. En otras palabras, es un

pequeño grupo que representa a toda la población. Y es así como las muestras probabilísticas surgen, pues también son consideradas como un subgrupo de la población, solo que en este caso todos los elementos tienen altas posibilidades de ser elegidos.

En esta investigación se utilizará una muestra del tipo probabilístico simple, por conveniencia propia, la cual estará conformado por un profesional por cada institución, encargado de la recopilación y difusión de la información de la biodiversidad de especies de mamíferos en la región San Martín, dicho grupo estará conformado por 12 personas.

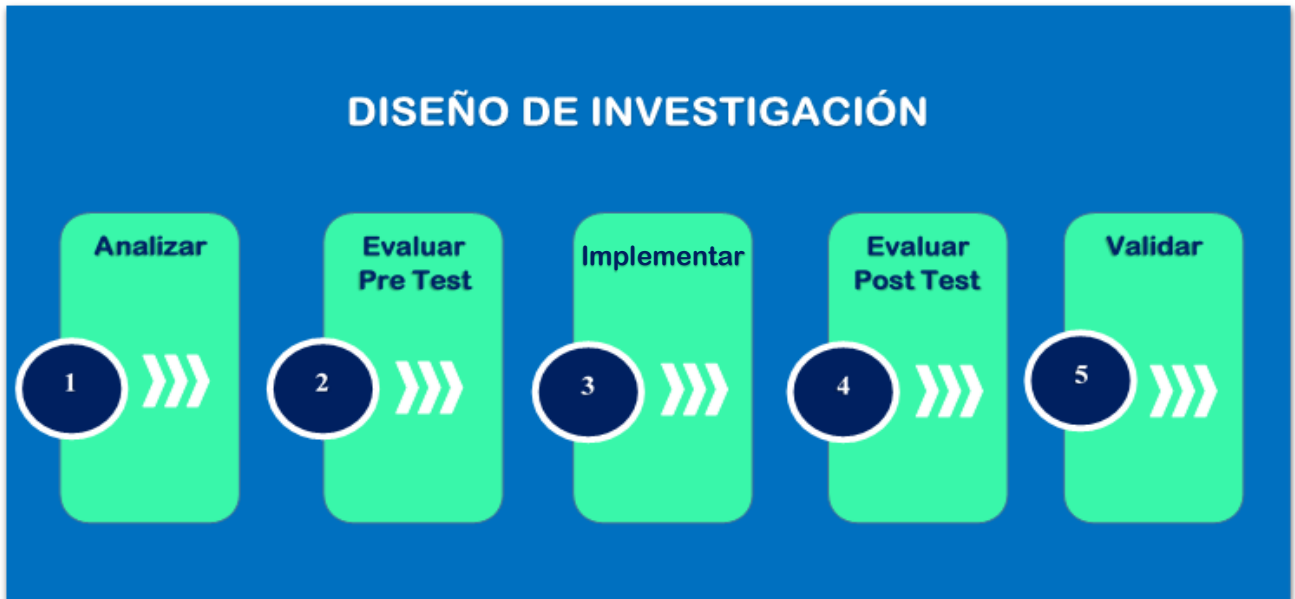
### **3.4. Diseño de investigación**

Según Salinas, (2012) considera al pre-experimento como una prueba no completada, es decir existe una medición hecha antes y después de ocurrido el efecto. A estos procedimientos se los llama Pre-test y Post-test, y es aplicado a un solo grupo.

Además, representa el corazón de la investigación, porque es allí donde se muestra la unión de todos los procedimientos que se llegaron a realizar como la descripción y planteamiento del problema, marco teórico y la hipótesis formulada. (Hernández et al., 2014)

Es por eso que para esta investigación consideraremos el diseño pre-experimental, debido a que evaluaremos al mismo grupo de interés en dos etapas. Antes y después de la implementación del buscador informático.

Ahora teniendo en cuenta la importancia del diseño de investigación. Describiremos las etapas que nos permitirán cumplir con nuestros objetivos. Las cuales están divididas en 4 fases:



*Figura 15 Diseño de investigación  
Fuente: Elaboración propia*

### **3.4.1. Analizar**

En esta etapa se pretende identificar la población objetivo y definir la muestra de estudio. Para cumplir con este objetivo, se pretende realizar visitas y entrevistas al Instituto de Investigaciones de la Amazonía Peruana.

### **3.4.2. Evaluar pretest**

Se tiene como objetivo evaluar al grupo de estudio, a través de una encuesta (pretest). Dicha evaluación nos permitirá conocer el estado actual del proceso de difusión de la información científica de las especies de mamíferos inventariados en la región San Martín.

### **3.4.3. Implementar**

Se entregará el buscador informático a los usuarios involucrados del Instituto de Investigaciones de la Amazonía Peruana, para que utilicen el sistema.

### **3.4.4. Evaluar posttest**

Luego según el cronograma establecido se procederá nuevamente a evaluar al grupo de estudio, a través de una posterior encuesta (posttest).

### **3.4.5. Validar**

Finalmente, llegaremos a comparar y validar los resultados obtenidos

## **3.5. Formulación de la hipótesis**

### **3.5.1. Hipótesis principal**

$H_1$ : El uso del Motor de Búsqueda basado en el Modelo de Espacio Vectorial mejorará la difusión de la información científicas de las especies de mamíferos inventariados en la Región San Martín.

$H_0$ : El uso del Motor de Búsqueda basado en el Modelo de Espacio Vectorial no mejorará la difusión de la información científicas de las especies de mamíferos inventariados en la Región San Martín.

### **3.5.2. Hipótesis específicas**

$H_1$ : El Motor de Búsqueda mejora la difusión de la información.

$H_0$ : El Motor de Búsqueda no mejora la difusión de la información.

$H_1$ : Identificar las actividades inherentes del proceso mejora la difusión de la información.

$H_0$ : Identificar las actividades inherentes del proceso no mejora la difusión de la información.

$H_1$ : Existe relación significativa entre el Motor de Búsqueda basado en el Modelo de Espacio Vectorial y la difusión de la información.

$H_0$ : No Existe relación significativa entre el Motor de Búsqueda basado en el Modelo de Espacio Vectorial y la difusión de la información.

## **3.6. Identificación de variables**



### 3.6.1. Matriz de consistencia

Tabla 3  
Matriz de consistencia

PROBLEMA	OBJETIVO	HIPOTESIS	VARIABLE	DIMENSIÓN	INDICADOR	INSTRUMENTO	METODOLOGÍA
<p><b>GENERAL:</b> ¿Cómo mejorar la difusión de la información científica de las especies de mamíferos inventariados en la región San Martín?</p>	<p><b>GENERAL:</b> Mejorar la difusión de la información científica de las especies de mamíferos inventariados en la región San Martín.</p>	<p><b>GENERAL:</b> El uso del Motor de Búsqueda basado en el Modelo de Espacio Vectorial mejorará la difusión de la información científica de las especies de mamíferos inventariados en la Región San Martín.</p>		<p><b>DIM.</b> Difusión (Castillo Blasco, 2014).</p>	<ul style="list-style-type: none"> <li>• Difusión bajo demanda</li> <li>• Difusión documentaria.</li> <li>• Medio de difusión utilizado</li> </ul>	<p><b>CUESTIONARIO</b></p>	<p><b>TIPO DE INVESTIGACIÓN:</b></p> <ul style="list-style-type: none"> <li>• Aplicada</li> </ul> <p><b>DISEÑO DE INVESTIGACIÓN:</b></p> <ul style="list-style-type: none"> <li>• Pre-experimental</li> </ul>
<p><b>ESPECÍFICOS:</b> ¿Cómo implementar el Motor de Búsqueda basado en el Modelo de Espacio Vectorial?</p>	<p><b>ESPECÍFICOS:</b> Implementar el Motor de Búsqueda basado en el Modelo de Espacio Vectorial</p>	<p><b>ESPECÍFICOS:</b> El Motor de Búsqueda mejora la difusión de la información</p>	<p><b>V.D:</b> Difusión de la información</p>				<p><b>POBLACIÓN Y MUESTRA:</b></p> <ul style="list-style-type: none"> <li>• El conjunto de personal de las 12 instituciones, responsables de la difusión de la información científica de la biodiversidad.</li> </ul> <p><b>TECNICA:</b></p> <ul style="list-style-type: none"> <li>• Encuesta</li> </ul>
<p>¿Cuál es la relación entre el Motor de Búsqueda basado en el Modelo de Espacio Vectorial y la difusión de la información</p>	<p>Determinar la relación de Motor de Búsqueda basado en el Modelo de Espacio Vectorial en la difusión de la información</p>	<p>Existe relación significativa entre el Motor de Búsqueda basado en el Modelo de Espacio Vectorial y la difusión de la información</p>					

---

¿Cuáles son las actividades inherentes a la difusión de la información?

Identificar las actividades inherentes a la difusión de la información

Las actividades inherentes del proceso mejoran la difusión de la información

**V.I:**  
Motor de Búsqueda basado en el Modelo de Espacio Vectorial

**DIM.**  
Sobre el motor Búsqueda

- Nivel de satisfacción.

### 3.6.2. Operacionalización de variables

Tabla 4

*Operacionalización de variables*

<b>Variable Independiente</b>	<b>Objetivos</b>	<b>Contenido</b>	<b>Método/ estrategia</b>	<b>Aplicación</b>
Motor de Búsqueda basado en el Modelo de Espacio Vectorial	Mejorar la difusión de la información científica de mamíferos inventariados en la región San Martín.		Modelo de recuperación vectorial	Se determinará si la implementación del buscador informático mejora la difusión de la información científica de biodiversidad
<b>Variable Dependiente</b>	<b>Dimensión</b>	<b>Indicadores</b>	<b>Def. Instrumentos</b>	<b>Def. Operacional</b>
Difusión de la información	Difusión	<ul style="list-style-type: none"> <li>Difusión bajo demanda (3p).</li> <li>Difusión documentaria (3p).</li> </ul>	<ul style="list-style-type: none"> <li>Considera que, en la región San Martín, la organización de la información científica de la biodiversidad de especies de mamíferos inventariados para su consulta es:</li> <li>La posibilidad de obtener una copia fiel de información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín es:</li> <li>La posibilidad de buscar información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín utilizando diferentes criterios (como nombre común, nombre científico, palabras clave, etc.), lo considera:</li> <li>La información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín que ofertan otras instituciones que manejan este tipo de información, lo considera:</li> <li>La modalidad (Motor de búsqueda) que se usa para transmitir la información científica de la biodiversidad de</li> </ul>	<ul style="list-style-type: none"> <li>Muy malo</li> <li>Malo</li> <li>Regular</li> <li>Bueno</li> <li>Muy bueno</li> </ul>

---

Sobre el motor de  
Búsqueda

- Medio de difusión utilizado (4p).
- Nivel de satisfacción (3p).

- especies de mamíferos inventariados de la región San Martín, lo considera:
- Cuando busca una información en particular, la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín sugeridos por otras instituciones que manejan este tipo de información, lo considera:
  - El libre acceso a la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín, lo considera:
  - La posibilidad de obtener toda la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín, lo considera:
  - El tiempo de respuesta en obtener la información requerida, lo considera:
  - La disponibilidad de la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín, lo considera:
  - Cuando solicita información científica sobre la biodiversidad de especies de mamíferos inventariados de la región San Martín en el motor de búsqueda ¿Cuál diría usted que es el nivel de similitud o coincidencia entre lo que usted requiere y los resultados obtenidos?
  - Cuando usted usa el buscador informático ¿Qué calificación le daría a la navegación intuitiva?
  - En cuanto a la sencillez de la interfaz ¿Qué calificación le daría usted a la interfaz del buscador informático?
-

### **3.7. Instrumento de recolección de datos**

El instrumento de medición a emplear es el cuestionario, la cual la aplicaremos a la muestra que forma parte del estudio. Y para su correcta aplicación es necesario tomar en cuenta “la validez y la confiabilidad, ya que estas no se asumen, se prueban”. (Hernández et al., 2014)

### **3.8. Técnica de recolección de datos**

Utilizaremos la técnica encuesta para la recolección de datos (con escala medida ordinal).

### **3.9. Validación de instrumentos**

La validación de los instrumentos se realizará mediante el juicio de expertos.

### **3.10. Plan de procesamiento de datos**

Para el procesamiento de datos analizaremos e interpretaremos los datos obtenidos del cuestionario con ayuda del software estadístico SPSS Statistics. También buscaremos dar explicación a los problemas planteados. Además, pretendemos realizar un contraste de normalidad, empleando la prueba de Shapiro Wilk, debido a que el tamaño de la muestra es menor a 50. Además, para contrastar hipótesis emplearemos la prueba paramétrica T-Student.

## **Capítulo 4**

### **CONSTRUCCIÓN DE LA PROPUESTA**

#### **4.1. Introducción**

En este capítulo mostraremos a detalle el proceso de construcción de la propuesta de solución, comenzando desde el análisis, evaluación de la situación problemática, e implementación del motor de búsqueda basado en el diseño de investigación.

#### **4.2. Analizar**

En esta etapa se realizó entrevistas al IIAP con el objetivo de obtener información sobre cómo funcionan a nivel organizacional, cuáles son sus prioridades, así mismo, para esto se realizó entrevistas al personal encargado del Programa de Investigación en Biodiversidad Amazónica – San Martín, que fue el programa de interés.

A continuación, se presenta algunos datos generales de interés que recolectamos, datos referentes a la misión y visión organizacional.

##### **4.2.1. Generalidades de la empresa**

###### **4.2.1.1. Misión**

Generar y proveer conocimientos sobre la diversidad biológica y socio-cultural de la Amazonía peruana, en beneficio de la población, que sean pertinentes, eficientes y confiables

###### **4.2.1.2. Visión**

Generar e incorporar conocimientos, tecnologías innovadoras y el saber ancestral, al servicio de las sociedades y de los ecosistemas amazónicos.

#### 4.2.1.3. *Objetivos estratégicos institucionales*

Tabla 5

*Objetivos estratégicos institucionales*

*Fuente: IIAP*

N° Objetivo estratégico	Descripción
OEI 1	Desarrollar los sistemas de producción sostenible en base a los recursos de la diversidad biológica amazónica utilizados por los productores.
OEI 2	Incrementar propuestas técnicas para la conservación y uso sostenible de la diversidad biológica, recursos hidrobiológicos, y bosques andino-amazónicos para uso de los órganos de desarrollo.

Para el desarrollo de la entrevista nos apoyamos de un conjunto de preguntas de interés elaboradas por el investigador. Ver *Figura 16* y *Figura 17*.

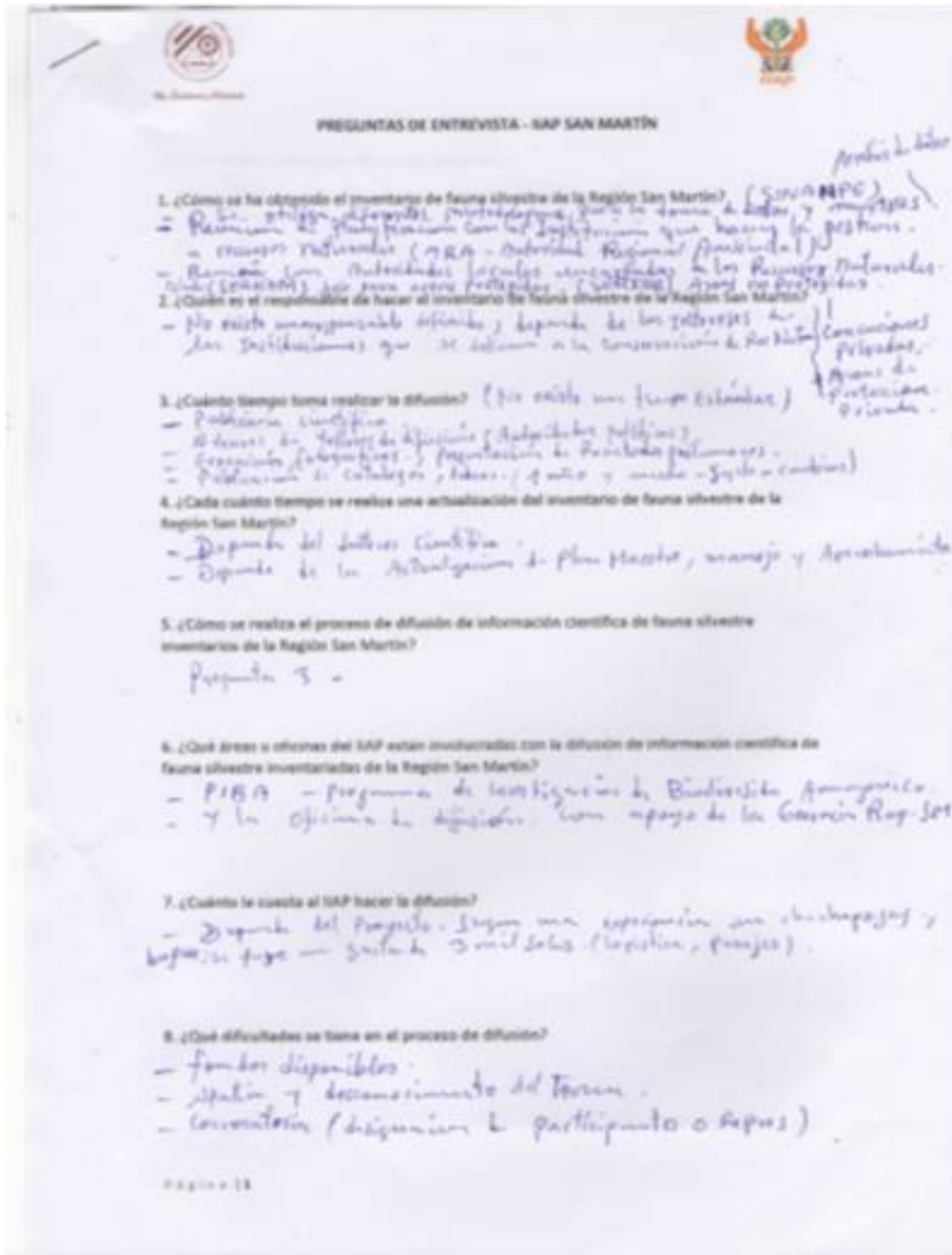


Figura 16 Preguntas de entrevista parte 1  
Fuente: Elaboración Propia



9. ¿Cuál es la población beneficiaria de estos datos?

- Todos (S M ) ; Población Vinculada a las Areas de Estudio .
- Depende los niveles ( Interesados ) -
- Comunidad Científica - ( Institutos, Universidad )
- Comunidad Académica ( Escuelas, Colegios )

10. ¿Considera que esa difusión es efectiva?, ¿Por qué?

- No, porq existe Apatía; se debe usar otros métodos.
- " , porq falta Recursos ,

11. ¿Qué etapas de la difusión considera que se pueden mejorar?, ¿Por qué?

- Recursos financieros ( Gestión )
- Convocatorias -
- Presupuesto .


Figura 17 Preguntas de entrevista parte 2  
Fuente: Elaboración Propia

### 4.3. Evaluar pretest

Luego del realizar visitas situacionales, y con el problema identificado, se procedió a aplicar un Pre-test al personal encargado de la difusión por cada institución del total de nuestra población de estudio. Sin embargo, debido a que la ubicación de algunas instituciones a evaluar es distante, nos apoyamos también de formulario google, ya que permite recopilar información de forma sencilla eficiente.



*Figura 18 Aplicando encuesta algunos representantes*



## ENCUESTA POS TEST

\*Obligatorio

Difusión Bajo Demanda

1. Considera que, en la región San Martín, la organización de la información científica de la biodiversidad de especies de mamíferos inventariados para su consulta es: \*

- Malo
- Regular
- Bueno
- Excelente

2. La posibilidad de obtener una copia fiel de información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín es: \*

- Malo
- Regular
- Bueno
- Excelente

3. La posibilidad de buscar información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín utilizando diferentes criterios (como nombre común, nombre científico, palabras clave, etc), lo considera: \*

- Malo
- Regular
- Bueno
- Excelente

ATRÁS

SIGUIENTE

Página 2 de 5

Este formulario se creó en Universidad Peruana Unión. [Notificar uso inadecuado](#) - [Condiciones del servicio](#)

Google Formularios

Figura 19 Pretest online de Google

#### **4.4. Implementar**

En esta fase se presenta el desarrollo de la construcción del motor de búsqueda, basado en una mixtura del modelo CRISP DM. Algunas actividades esenciales fueron la configuración de la plantilla del sitio web, ver *Anexos 12*, así mismo tuvo lugar la organización y modelamiento de los datos, como también la materialización de los vectores dentro de las consultas de información. A continuación, presentamos parte de este proceso que va desde la comprensión de los datos hasta la obtención de un ranking de resultados:

##### **4.4.1. Comprensión de los datos**

En esta etapa, se procedió a organizar la data de los inventarios de los mamíferos de la región, esta data se logró obtener gracias a la colaboración del IIAP San Martín.

Utilizamos la hoja de cálculo de Microsoft Excel, considerando los siguientes campos: nombre científico, reino, filo, clase, orden, familia, género, sinónimo, nombre común, autor del nombre científico, ubicación, referencia bibliográfica y url de la imagen, producto de esto obtuvimos una base de datos en Excel. Ver *Figura 20*.

	Nombre científico	Reino	Filo	Clase	Orden	Familia	Genero	Sinonimos	Nombre común	Autor del Nombre científica	Ubicación	Referencia bibliográfica	URL Imagen
4	<i>Ateles paniscus</i>	Animalia	Chordata	Mammalia	Primates	Aleidae	Ateles		Mono araña negro	Linnaeus, 1758		ERDB-SM	<a href="http://www.arkive.org/black-spider-monkey/atesles-paniscus/#src=portletV3api">http://www.arkive.org/black-spider-monkey/atesles-paniscus/#src=portletV3api</a>
5	<i>Saguinus spp</i>	Animalia	Chordata	Mammalia	Primates	Callitrichidae	Saguinus		pichicos	Hoffmannsegg, 1807		ERDB-SM	
6	<i>Saguinus mystax</i>	Animalia	Chordata	Mammalia	Primates	Callitrichidae	Saguinus		Pichico barba blanca	Spix, 1803		ERDB-SM	<a href="http://www.arkive.org/moustached-tamarin/saguinus-mystax/#src=portletV3api">http://www.arkive.org/moustached-tamarin/saguinus-mystax/#src=portletV3api</a>
8	<i>Choleopus sp.</i>	Animalia	Chordata	Mammalia	Pilosa	Megalonychidae	Choleopus		Peresozoo	Illiger, 1811		PM-ACROE	
9	<i>Agouti paca</i>	Animalia	Chordata	Mammalia	Rodentia	Agoutidae	Agouti		Majaz	Linnaeus, 1766		ERDB-SM, PM-ACROE, PM-ACMAHARAM, PM-BPAM	<a href="http://3.bp.blogspot.com/_opp4FMMUoTjGJTjIENAAI/AAAAAAACJ4I23k8KSVuJ8Qis1600/abba%2520-%2520paca-2">http://3.bp.blogspot.com/_opp4FMMUoTjGJTjIENAAI/AAAAAAACJ4I23k8KSVuJ8Qis1600/abba%2520-%2520paca-2</a>
10	<i>Akodon aerosus</i>	Animalia	Chordata	Mammalia	Rodentia	Muridae	Akodon		Highland Grass Mouse	Thomas, 1913		PM-ACROE	<a href="http://zoologia.puce.edu.ec/Vertebrados/mamiferos/FichaEspecie.aspx?Id=2095">http://zoologia.puce.edu.ec/Vertebrados/mamiferos/FichaEspecie.aspx?Id=2095</a>
11	<i>Alouatta seniculus</i>	Animalia	Chordata	Mammalia	Primates	Aleidae	Alouatta		Coto mono, mono aullador	Linnaeus, 1766		PM-ACROE, PM-ACROE, PM-ACMAHARAM, PM-PNRA, PM-BPAM, FJDES, 2005	<a href="http://www.arkive.org/colombian-red-howler-monkey/alouatta-seniculus/#src=portletV3api">http://www.arkive.org/colombian-red-howler-monkey/alouatta-seniculus/#src=portletV3api</a>
12	<i>Anoura sp.</i>	Animalia	Chordata	Mammalia	Chiroptera	Phyllostomidae	Anoura			Gray, 1838		PM-ACROE	
13	<i>Atotus brumbacki</i>	Animalia	Chordata	Mammalia	Primates	Aotidae	Atotus		Musmuqui	Hershkovitz, 1983		PM-PNRA	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FPrimates%2F%2FCebide%2FAotuBru1.jpg">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FPrimates%2F%2FCebide%2FAotuBru1.jpg</a>
14	<i>Atotus Lemurinus</i>	Animalia	Chordata	Mammalia	Primates	Aotidae	Atotus		Musmuqui	I. Geoffroy, 1846		PM-PNRA	<a href="http://cmap.upb.edu.co/nd=1195239996390_1210733249_59459_Aotus%20lemurinus.jpg">http://cmap.upb.edu.co/nd=1195239996390_1210733249_59459_Aotus%20lemurinus.jpg</a>
15	<i>Atotus miconax</i>	Animalia	Chordata	Mammalia	Primates	Aotidae	Atotus		Musmuqui andino, tocon del mayo	Thomas, 1927		PM-BPAM	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FPrimates%2F%2FCebide%2FAotuMic1.jpg">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FPrimates%2F%2FCebide%2FAotuMic1.jpg</a>
16	<i>Atotus nigriceps</i>	Animalia	Chordata	Mammalia	Primates	Aotidae	Atotus		Tutamono	Dollman, 1909		PM-ACMAHARAM	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FPrimates%2F%2FCebide%2FAotuNig1.jpg">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FPrimates%2F%2FCebide%2FAotuNig1.jpg</a>
17	<i>Atotus sp.</i>	Animalia	Chordata	Mammalia	Primates	Aotidae	Atotus		musmuqui	Spix, 1823		PM-ACROE	
18	<i>Atotus trivirgatus</i>	Animalia	Chordata	Mammalia	Primates	Aotidae	Atotus		Musmuqui	Humboldt, 1812		PM-PNRA	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FPrimates%2F%2FCebide%2FAotuTr3.jpg">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FPrimates%2F%2FCebide%2FAotuTr3.jpg</a>
19	<i>Atotus vociferans</i>	Animalia	Chordata	Mammalia	Primates	Cebidae	Atotus		Marikón Grillo, Musmuqui	Spix, 1823		PM-ACMAHARAM, PM-ACROE, PM-PNRA	<a href="http://www.arkive.org/night-monkey/atotus-vociferans/#src=portletV3api">http://www.arkive.org/night-monkey/atotus-vociferans/#src=portletV3api</a>
20	<i>Artibeus planirostris</i>	Animalia	Chordata	Mammalia	Chiroptera	Phyllostomidae	Artibeus		Fiat-faced Fruit-eating Bat	Spix, 1823		PM-ACROE	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FVolants%2FPhyllost%2FArtPla3.jpg">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FVolants%2FPhyllost%2FArtPla3.jpg</a>
21	<i>Artibeus lituratus</i>	Animalia	Chordata	Mammalia	Chiroptera	Phyllostomidae	Artibeus		Great Fruit-eating Bat	Olfers, 1818		PM-ACROE	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FVolants%2FPhyllost%2FArtLit5.jpg">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FVolants%2FPhyllost%2FArtLit5.jpg</a>
22	<i>Artibeus obscurus</i>	Animalia	Chordata	Mammalia	Chiroptera	Phyllostomidae	Artibeus		Dark Fruit-eating Bat	Schinz, 1821		PM-ACROE	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FVolants%2FPhyllost%2FArtObs2.jpg">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FVolants%2FPhyllost%2FArtObs2.jpg</a>
23	<i>Ateles belzebuth</i>	Animalia	Chordata	Mammalia	Primates	Cebidae	Ateles		Ateo Peludo	É. Geoffroy, 1806		PM-ACROE, propuesta ACAMM Allo Shunib-Toache, PM-PNRA, PM-BPAM	<a href="http://www.arkive.org/white-bellied-spider-monkey/atesles-belzebuth/#src=portletV3api">http://www.arkive.org/white-bellied-spider-monkey/atesles-belzebuth/#src=portletV3api</a>
24	<i>Ateleocynus microtis</i>	Animalia	Chordata	Mammalia	Carnivora	Canidae	Ateleocynus		Zorro de oreja corta, sachapero	Solater, 1883		PM-ACROE, PM-ACMAHARAM, PM-PNRA	<a href="http://www.arkive.org/small-eared-zorro/ateleocynus-microtis/#src=portletV3api">http://www.arkive.org/small-eared-zorro/ateleocynus-microtis/#src=portletV3api</a>
25	<i>Bradypus tridactylus</i>	Animalia	Chordata	Mammalia	Pilosa	Bradypodidae	Bradypus		Oso perezoso, chovejo	Linnaeus, 1758		PM-ACMAHARAM	<a href="http://www.arkive.org/pale-throated-three-toed-sloth/bradypus-tridactylus/">http://www.arkive.org/pale-throated-three-toed-sloth/bradypus-tridactylus/</a>
26	<i>Bradypus variegatus</i>	Animalia	Chordata	Mammalia	Xenarthra	Bradypodidae	Bradypus		Peresozo de tres dedos	Schinz, 1825		PM-ACROE	<a href="http://www.arkive.org/brown-throated-three-toed-sloth/bradypus-variegatus/">http://www.arkive.org/brown-throated-three-toed-sloth/bradypus-variegatus/</a>
27	<i>Bradypus sp.</i>	Animalia	Chordata	Mammalia	Pilosa	Bradypodidae	Bradypus		pelejo/perezoso	Schinz, 1825		PM-ACROE	
28	<i>Callicebus oenanthe</i>	Animalia	Chordata	Mammalia	Primates	Pitheciidae	Callicebus		Tocon andino	Thomas, 1924		PM-BPAM	<a href="http://www.arkive.org/san-marín-tin-monkey/callicebus-oenanthe/">http://www.arkive.org/san-marín-tin-monkey/callicebus-oenanthe/</a>
29	<i>Carollia brevicauda</i>	Animalia	Chordata	Mammalia	Chiroptera	Carollidae	Carollia		Silky Short-tailed Bat	Schinz, 1821		PM-ACROE	<a href="http://www.arkive.org/silky-short-tailed-bat/carollia-brevicauda/">http://www.arkive.org/silky-short-tailed-bat/carollia-brevicauda/</a>
30	<i>Carollia castanea</i>	Animalia	Chordata	Mammalia	Chiroptera	Carollidae	Carollia		Chestnut Short-tailed Bat	Allen, 1890		PM-ACROE	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FVolants%2FPhyllost%2FCarCas2.jpg">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FVolants%2FPhyllost%2FCarCas2.jpg</a>
31	<i>Carollia perspicillata</i>	Animalia	Chordata	Mammalia	Chiroptera	Carollidae	Carollia		Seb's Short-tailed Bat	Linnaeus, 1758		PM-ACROE	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FVolants%2FPhyllost%2FCarPer2.jpg">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FVolants%2FPhyllost%2FCarPer2.jpg</a>
32	<i>Cebuella pygmaea</i>	Animalia	Chordata	Mammalia	Primates	Callitrichidae	Cebuella		Leoncito	Spix, 1823		PM-ACMAHARAM, PM-ACROE, PM-PNRA	<a href="http://www.arkive.org/bygnny-marmoset/cebuella-pygmaea/#src=portletV3api">http://www.arkive.org/bygnny-marmoset/cebuella-pygmaea/#src=portletV3api</a>
33	<i>Cebus apella</i>	Animalia	Chordata	Mammalia	Primates	Cebidae	Cebus		Capuchino de Cabeza Dura	Linnaeus, 1758		PM-ACROE, PM-ACMAHARAM, PM-PNRA	<a href="http://www.arkive.org/black-capped-capuchin/cebus-apella/">http://www.arkive.org/black-capped-capuchin/cebus-apella/</a>
34	<i>Cebus albifrons</i>	Animalia	Chordata	Mammalia	Primates	Cebidae	Cebus		Machin blanco	Humboldt, 1812		ERDB-SM, PM-ACROE, PM-PNRA	<a href="http://www.arkive.org/white-fronted-capuchin/cebus-albifrons/">http://www.arkive.org/white-fronted-capuchin/cebus-albifrons/</a>
35	<i>Chiroderma trinitatum</i>	Animalia	Chordata	Mammalia	Chiroptera	Carollidae	Chiroderma		Little Big-eyed Bat	Goodwin, 1958		PM-ACROE	<a href="http://www.arkive.org/guadeloupean-big-eyed-bat/chiroderma-improvisum/">http://www.arkive.org/guadeloupean-big-eyed-bat/chiroderma-improvisum/</a>
36	<i>Chironectes minimus</i>	Animalia	Chordata	Mammalia	Didelphimorphia	Didelphidae	Chironectes		Señuela de agua, pintonaman	Zimmermann, 1780		PM-ACROE, PM-ACMAHARAM	<a href="http://www.arkive.org/water- opossum/chironectes-minimus/">http://www.arkive.org/water- opossum/chironectes-minimus/</a>
37	<i>Choleopus didactylus</i>	Animalia	Chordata	Mammalia	Pilosa	Megalonychidae	Choleopus		pelejo	Linnaeus, 1758		PM-ACMAHARAM	<a href="http://www.arkive.org/southern-two-toed-sloth/choleopus-didactylus/">http://www.arkive.org/southern-two-toed-sloth/choleopus-didactylus/</a>
38	<i>Coendou prehensilis</i>	Animalia	Chordata	Mammalia	Rodentia	Erethizontidae	Coendou		Puercoepoin arbicola	Linnaeus, 1758		PM-ACROE	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FRongeur%2FCaviomo%2FPorcEpic%2FCoen">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FRongeur%2FCaviomo%2FPorcEpic%2FCoen</a>
39	<i>Coendou bicolor</i>	Animalia	Chordata	Mammalia	Rodentia	Erethizontidae	Coendou		Puercoepoin, cachacushillo	Tschudi, 1844		PM-ACROE, PM-ACMAHARAM	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FRongeur%2FCaviomo%2FPorcEpic%2FCoen">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FRongeur%2FCaviomo%2FPorcEpic%2FCoen</a>
40	<i>Conospatus semistriatus</i>	Animalia	Chordata	Mammalia	Carnivora	Mephitidae	Conospatus		Zorrillo	Boddaert, 1785		PM-ACROE, PM-ACMAHARAM	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FCarniv%2FMustelid%2FConeSem4.jpg">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FCarniv%2FMustelid%2FConeSem4.jpg</a>
41	<i>Cuniculus paca</i>	Animalia	Chordata	Mammalia	Rodentia	Dasyproctidae	Cuniculus		Paca	Linnaeus, 1766		PM-ACMAHARAM, PM-ACROE	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FRongeur%2FCaviomo%2FAu%2FCav%2FAgu">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FRongeur%2FCaviomo%2FAu%2FCav%2FAgu</a>
42	<i>Cyclopes didactylus</i>	Animalia	Chordata	Mammalia	Pilosa	Cyclopedidae	Cyclopes		Serafin, oso homiguero	Linnaeus, 1758		PM-ACROE, PM-ACMAHARAM	<a href="http://www.arkive.org/bygnny-anteater/cyclopes-didactylus/#src=portletV3api">http://www.arkive.org/bygnny-anteater/cyclopes-didactylus/#src=portletV3api</a>
43	<i>Dasyprocta fuliginosa</i>	Animalia	Chordata	Mammalia	Rodentia	Dasyproctidae	Dasyprocta		Añuje marrón	Wagner, 1832		PM-ACROE, PM-ACMAHARAM, PM-BPAM	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FRongeur%2FCaviomo%2FAu%2FCav%2FDasy">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FRongeur%2FCaviomo%2FAu%2FCav%2FDasy</a>
44	<i>Dasyprocta variegata</i>	Animalia	Chordata	Mammalia	Rodentia	Dasyproctidae	Dasyprocta		Añuje gris	Goldman, 1913		PM-ACROE, ERDB-SM	<a href="http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FRongeur%2FCaviomo%2FAu%2FCav%2FDasy">http://www.planet-mammiferes.org/dupal/en/node/40?indice2=Photos%2FRongeur%2FCaviomo%2FAu%2FCav%2FDasy</a>
45	<i>Dasyopus pilosus</i>	Animalia	Chordata	Mammalia	Cingulata	Dasyopidae	Dasyopus		Armadillo peludo	Fitzinger, 1856		PM-PNRA	<a href="http://www.lacredlist.org/details/links/6291/0">http://www.lacredlist.org/details/links/6291/0</a>
46	<i>Dasyopus novemcinctus</i>	Animalia	Chordata	Mammalia	Cingulata	Dasyopidae	Dasyopus		Carachupa, armadillo	Linnaeus, 1758		ERDB-SM, PM-ACMAHARAM, PM-ACROE, PM-BPAM	<a href="http://www.arkive.org/nine-banded-armadillo/dasyopus-novemcinctus/#src=portletV3api">http://www.arkive.org/nine-banded-armadillo/dasyopus-novemcinctus/#src=portletV3api</a>
47	<i>Desmodus rotundus</i>	Animalia	Chordata	Mammalia	Chiroptera	Phyllostomidae	Desmodus		Common Vampire Bat	É. Geoffroy, 1810		PM-ACROE	<a href="http://www.arkive.org/common-vampire-bat/desmodus-rotundus/">http://www.arkive.org/common-vampire-bat/desmodus-rotundus/</a>

Figura 20 Base de datos en Excel  
Fuente: Elaboración propia

#### 4.4.2. Modelado de los datos

Para el modelado de nuestra base de datos utilizamos la plataforma de MongoDB ya que nos permite manejar un esquema dinámico, habiendo que la integración de los datos sea más sencilla.

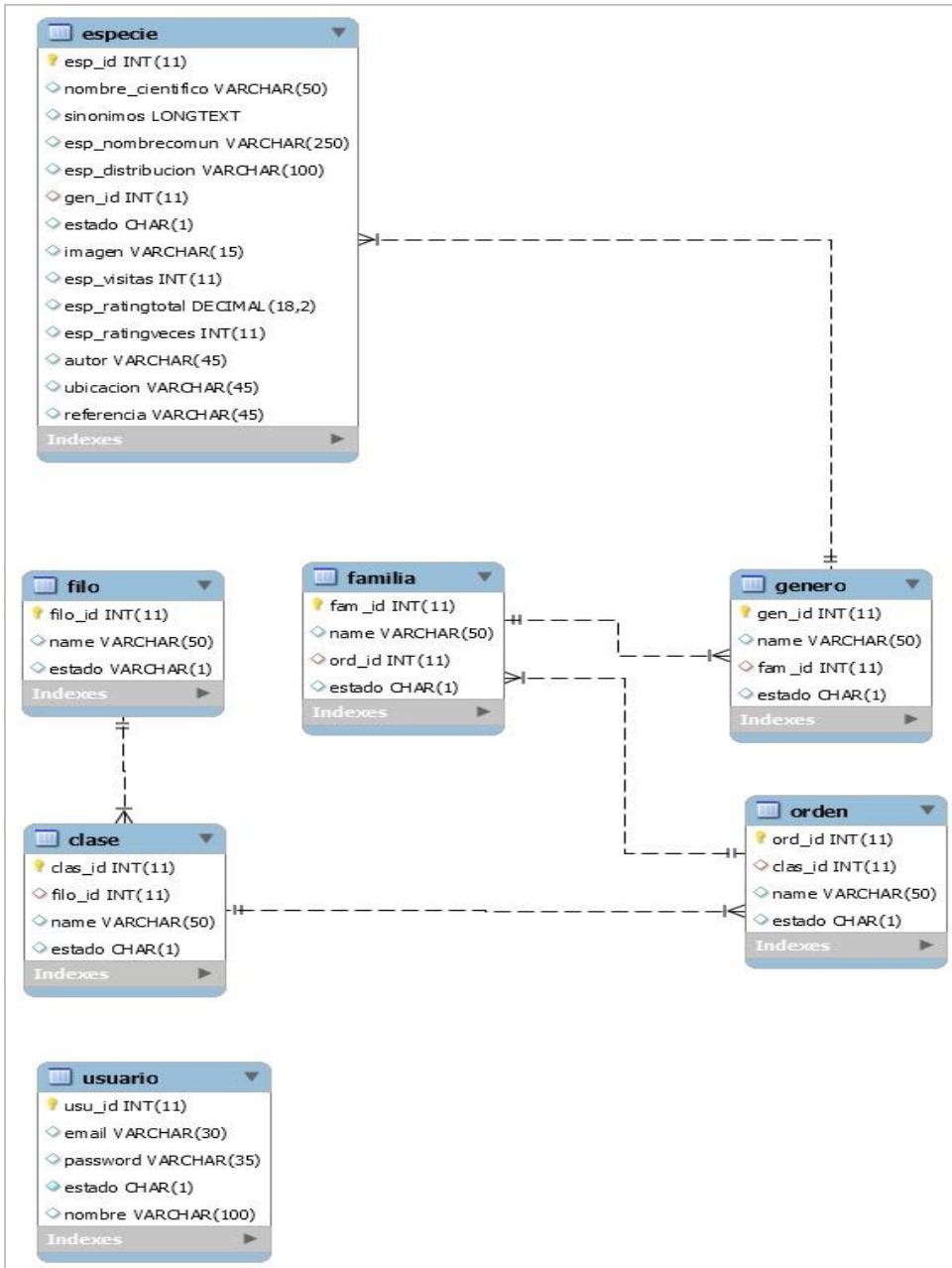


Figura 21 Base de datos del sistema

### 4.4.3. Evaluación

En esta etapa se evaluó los resultados obtenidos a partir del funcionamiento del buscador, específicamente en el tiempo requerido para realizar la búsqueda. Así mismo como la funcionalidad interna del ingreso, edición y modificación de los datos.

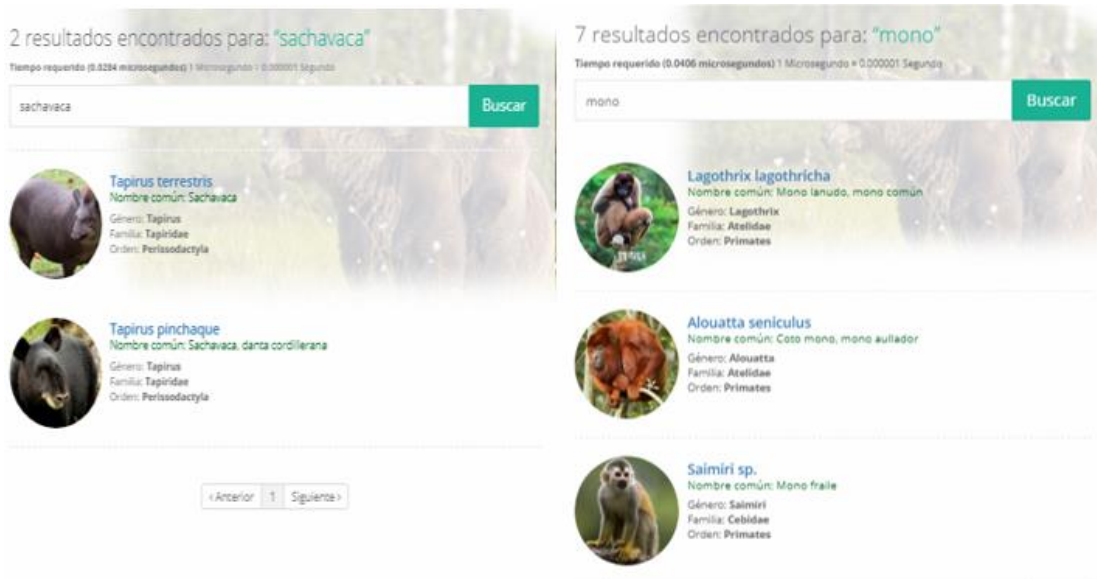


Figura 22 Evaluando resultados del MB -1



Figura 23 Evaluando resultados del MB -2

En la *Figura 22* y *Figura 23* los resultados del motor de búsqueda se muestran favorables en cuanto a la capacidad de su pronta respuesta. Algunas cifras son de 0.0232 microsegundos de tiempo requerido para realizar una búsqueda. Así mismo, en la *Figura 25* y *Figura 26*, los pruebas realizadas en cuanto al registro, edición y modificación de los datos, lo cual se obtuvo resultados favorables.



Figura 24 Tiempo requerido para realizar búsqueda



Editar Especie


<b>Género</b>	Mazama <input type="button" value="Q"/>	<b>Autor (Opcional)</b>	Rafinesque, 1817
<b>Nombre Científico</b>	<input type="text" value="Mazama sp."/>	<b>Ubicación (Opcional)</b>	
<b>Sinónimos (Opcional)</b>	<input type="text"/>	<b>Referencia (Opcional)</b>	PM-ACMAHARAM
<b>Nombre común (Opcional)</b>	<input type="text" value="venado"/>		
<b>Imagen</b>	 <input type="button" value="Seleccionar archivo"/> Ningún archiv...seleccionado		

Figura 25 Evaluando la función editar

**Agregar Especie**

Género: Leopardus

Nombre Científico: Panthera pardus

Sinónimos (Opcional):

Nombre común (Opcional): Leopardo

Imagen: Seleccionar archivo Ningún archiv...seleccionado

Autor (Opcional):

Ubicación (Opcional):

Referencia (Opcional):

Cerrar Guardar

Figura 26 Evaluando la función agregar

**Especies de Mamíferos**

Especies / Especies de Mamíferos

+ Agregar Editar Eliminar

Mostrar: 10

Buscar:

Código	Género
201	Vampyrodes
200	Vampyriscus
199	Vampyressa
198	Uroderma
197	Tremarctos
196	Tonatia
195	Thomasomys
194	Tayassu
193	Tapirus
192	Tapirus

Mostrando registros del 1 al 10 de un total de 123 registros 1 row selected

Anterior 1 2 3 4 5 ... 13

Figura 27 Evaluando la función eliminar

#### **4.4.4. Implantación**

Luego de haber realizado algunas pruebas al buscador, presentamos el proceso de transformación de los documentos en vectores que serán de utilidad para el motor de búsqueda al momento de recomendar al usuario que resultado se asemeja a su consulta. Es necesario recordar lo que dijimos al principio, en el modelo de espacio vectorial se emplea los pesos de los términos para cada documento (TD), que refleja la relevancia o significancia de los TD al momento de presentarse la colección.

A continuación, presentamos las etapas de este proceso que va desde cuando el usuario realiza la consulta hasta la obtención de un ranking de resultados, que tienen un mayor grado de similitud con la consulta.

#### **4.4.5. Realizar Consulta**

Tomando como nuestro primer punto de referencia, se presenta en la siguiente figura una actividad común de un usuario intentando obtener respuestas a su consulta, y sobre todo poder encontrar información relevante y confiable, también se puede ver un ranking de resultados similares a su consulta.


← → ↻ ⓘ No es seguro layza.infinitemperu.com/results?options=aves&query=Mono 🔍 ☆ ABP 📧 🗑️ 🏠

🏠 Regresar a Principal Inventario de Mamíferos de la Región San Martín Admin 🗨️

7 resultados encontrados para: "Mono"  
 Tiempo requerido (0.6674 microsegundos) 1 Microsegundo = 0.000001 Segundo


Mono Buscar

---




**Lagothrix lagothericha**  
 Nombre común: Mono lanudo, mono común  
 Género: Lagothrix  
 Familia: Atellidae  
 Orden: Primates

---




**Alouatta seniculus**  
 Nombre común: Coto mono, mono aullador  
 Género: Alouatta  
 Familia: Atellidae  
 Orden: Primates

---




**Saimiri sp.**  
 Nombre común: Mono fraile  
 Género: Saimiri  
 Familia: Cebidae  
 Orden: Primates

---




**Saimiri sciureus**  
 Nombre común: Mono ardilla, frailecillo  
 Género: Saimiri  
 Familia: Cebidae  
 Orden: Primates

---



**Ateles paniscus**  
 Nombre común: Mono araña negro  
 Género: Ateles  
 Familia: Atellidae  
 Orden: Primates

---

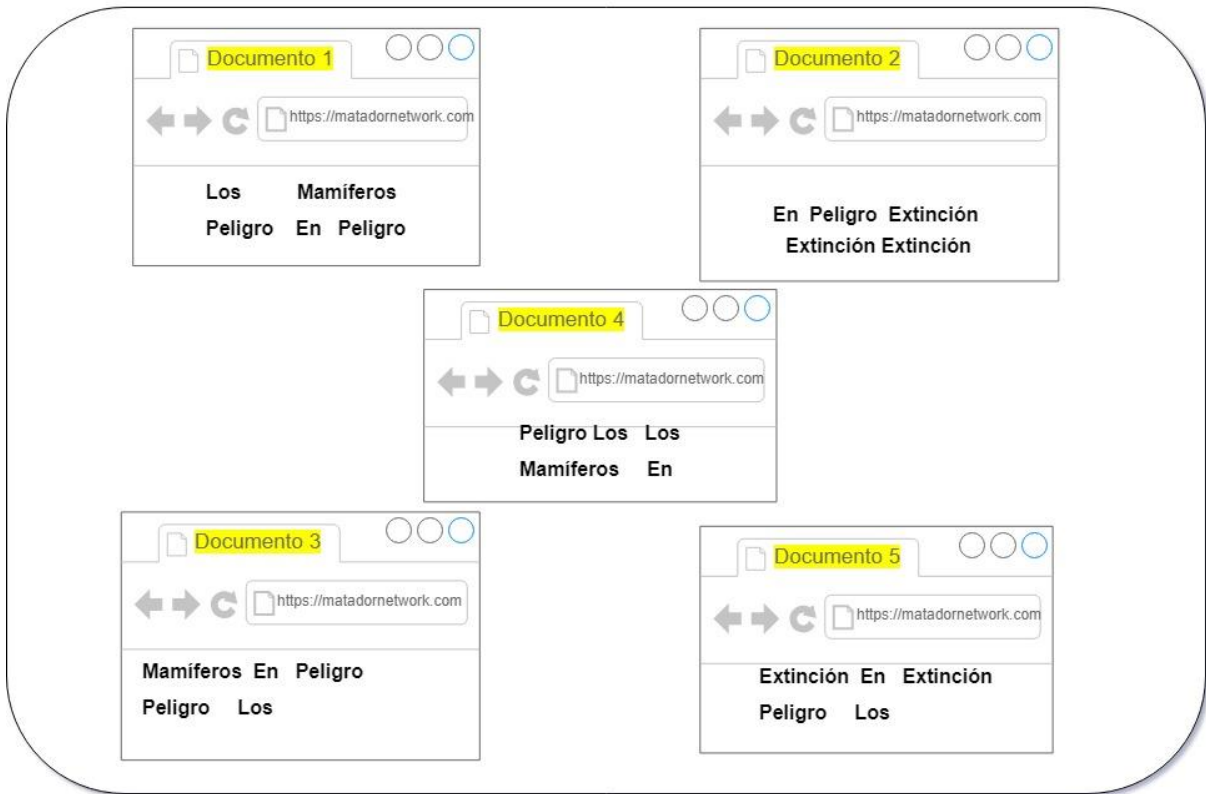


**Saguinus fuscicollis**  
 Nombre común: Tamarino de Cabeza Amarilla, mono pichico  
 Género: Saguinus  
 Familia: Callitrichidae  
 Orden: Primates

*Figura 28 Consulta de un usuario*

#### 4.4.6. Representación de la colección de documentos

La *Figura 29* muestra un ejemplo de un esquema de la colección de documentos. Esta segunda etapa consiste en organizar los documentos a partir de las palabras claves que más veces se repiten en el mismo.



*Figura 29 Representación de la colección de documentos*

#### 4.4.7. Representación de una lista invertida

Como tercer paso la *Tabla 6* muestra una representación de la lista invertida de los términos, donde en la matriz (1,1) el primer número representa el número de documento, y el segundo número la suma de todas las ocurrencias ó el número de veces que aparece un término en un documento.

Tabla 6  
Representación final del vocabulario

N°	Vocabulario	Lista invertida	IDF
1	Los	(1,1) (3,1) (4,2) (5,1)	$\text{Log}(5/4)$
2	Mamíferos	(1,1) (3,1) (4,1)	$\text{Log}(5/3)$
3	Peligro	(1,2) (2,1) (3,2) (4,1) (5,1)	$\text{Log}(5/5)$
4	En	(1,1) (2,1) (3,1) (4,1) (5,1)	$\text{Log}(5/5)$
5	Extinción	(2,3) (5,2)	$\text{Log}(5/2)$

Para una mejor explicación tomaremos la palabra mamíferos. La palabra aparece en el documento 1 sólo 1 vez, así mismo, en el documento 3 aparece 1 sola vez, y finalmente en el documento 4 aparece solo 1 vez. Del mismo modo aplica para los demás términos.

#### 4.4.8. Cálculo del IDF a partir de la lista invertida

Luego de obtener la lista invertida, aplicamos la fórmula del IDF, ver *Figura 5*. El cálculo consiste en aplicar el logaritmo del número total de documentos de la colección, dividido entre el número de documentos en los que aparece el término a lo largo de toda la colección.

Tabla 7  
Lista invertida

	Vocabulario	Lista invertida	IDF
1	Los	(1,1) (3,1) (4,2) (5,1)	0.32192809
2	Mamíferos	(1,1) (3,1) (4,1)	0.73696559
3	Peligro	(1,2) (2,1) (3,2) (4,1) (5,1)	0
4	En	(1,1) (2,1) (3,1) (4,1) (5,1)	0
5	Extinción	(2,3) (5,2)	1.32192809
6	Q	Mamífero en peligro peligro	

La *Tabla 7* muestra el resultado de la operación de la fórmula del IDF (ver *Figura 5*), así mismo se añadió la consulta del usuario al final de la columna de la lista invertida.

#### 4.4.9. Ponderación TF – IDF

La siguiente *Tabla 8* muestra el resultado de la ponderación de los términos, es decir los pesos de las palabras. La operación consiste en multiplicar la frecuencia de aparición de un término (TF) en un documento, por el factor IDF de un término. Ver *Figura 6*. Del mismo modo se efectúa la operación para la consulta “*Mamíferos en peligro peligro*”.

Tabla 8  
*Ponderación de los términos*

TF *	Los	Mamíferos	Peligro	En	Extinción
IDF					
d1	0.3219280	0.73696559	0	0	0
d2	0	0.73696559	0	0	3.965784
d3	0.3219280	0.73696559	0	0	0
d4	0.64385619	0.73696559	0	0	0
d5	0	0	0	0	2.6438562
Q	0	0.73696559	0	0	0

Para una mejor explicación de la operación, tomaremos la palabra “los”. El cálculo parte tomando el IDF de la palabra “los” (0.32192809), multiplicado por el TF del mismo en el d1, como “los” solo aparece 1 sola vez en el documento 1, el resultado es la misma cantidad (0.32192809), y así del mismo modo para el resto de los términos.



#### 4.4.10. Cálculo de similitud del documento con la consulta

Tabla 9

*Cálculo de Similitud del documento con la consulta*

	Operación	Resultad
		o
Sim d1, Q	$\frac{(0.32 * 0) + (0.73 * 0.73) + (0 * 0) + (0 * 0) + (0 * 0)}{\sqrt{(0.32)^2 + (0.73)^2 + (0.0)^2 + (0.0)^2 + (0.0)^2} * \sqrt{(0.0)^2 + (0.73)^2 + (0.0)^2 + (0.0)^2 + (0.0)^2}}$	0.49770
Sim d2, Q	$\frac{(0 * 0) + (0.73 * 0.73) + (0 * 0) + (0 * 0) + (3.96 * 0)}{\sqrt{(0)^2 + (0.73)^2 + (0.0)^2 + (0.0)^2 + (3.96)^2} * \sqrt{(0.0)^2 + (0.73)^2 + (0.0)^2 + (0.0)^2 + (0.0)^2}}$	0.09922
Sim d3, Q	$\frac{(0.32 * 0) + (0.73 * 0.73) + (0 * 0) + (0 * 0) + (0 * 0)}{\sqrt{(0.32)^2 + (0.73)^2 + (0.0)^2 + (0.0)^2 + (0.0)^2} * \sqrt{(0.0)^2 + (0.73)^2 + (0.0)^2 + (0.0)^2 + (0.0)^2}}$	0.49770
Sim d4, Q	$\frac{(0.64 * 0) + (0.73 * 0.73) + (0 * 0) + (0 * 0) + (0 * 0)}{\sqrt{(0.64)^2 + (0.73)^2 + (0.0)^2 + (0.0)^2 + (0.0)^2} * \sqrt{(0.0)^2 + (0.73)^2 + (0.0)^2 + (0.0)^2 + (0.0)^2}}$	0.40900
Sim d5, Q	$\frac{(0.32 * 0) + (0 * 0.73) + (0 * 0) + (0 * 0) + (2.64 * 0)}{\sqrt{(0.32)^2 + (0)^2 + (0.0)^2 + (0.0)^2 + (2.64)^2} * \sqrt{(0.0)^2 + (0.73)^2 + (0.0)^2 + (0.0)^2 + (0.0)^2}}$	0.0

Como penúltimo paso se procede a operar con la siguiente formula:

$$\text{Cos } \theta = \frac{(\vec{A}_{\text{los}} * \vec{B}_{\text{los}}) + (\vec{A}_{\text{mamíferos}} * \vec{B}_{\text{mamíferos}}) + (\vec{A}_{\text{peligro}} * \vec{B}_{\text{peligro}}) + (\vec{A}_{\text{en}} * \vec{B}_{\text{en}}) + (\vec{A}_{\text{extinción}} * \vec{B}_{\text{extinción}})}{|\vec{A}| * |\vec{B}|}$$

Donde:

$\vec{A}_{\text{los}}$ : Peso del término en el d1

$\vec{B}_{\text{los}}$ : Peso de la consulta en d1

$\vec{A}_{\text{mamíferos}}$ : Peso del término en el d1

$\vec{B}_{\text{mamíferos}}$ : Peso de la consulta en d1

$\vec{A}_{\text{peligro}}$ : Peso del término en el d1

$\vec{B}_{\text{peligro}}$ : Peso de la consulta en d1

$\vec{A}_{\text{en}}$ : Peso del término en el d1

$\vec{B}_{\text{en}}$ : Peso de la consulta en d1

$\vec{A}_{\text{extinción}}$ : Peso del término en el d1

$\vec{B}_{\text{extinción}}$ : Peso de la consulta en d1

Los resultados obtenidos nos permitirán tener los pesos finales de cada término, para luego después elaborar el ranking.

#### **4.4.11. Ranking de consulta**

Por último considerando la consulta “Mamíferos en peligro peligro”, en la *Tabla 10* se presenta un ranking los resultados obtenidos, esto es fruto del cálculo de similaridad, es decir se considera al documento cuyo resultado del peso sea mayor que el de todos, y es así de forma descendente tenemos el ranking de los resultados.

Tabla 10  
*Ranking de la consulta*

Documento	Términos del documento				
D3	mamíferos	en	peligro	peligro	los
D1	los	mamíferos	peligro	peligro	en
D4	peligro	los	los	mamíferos	en
D2	en	peligro	extinción	extinción	extinción
D5	extinción	en	extinción	peligro	los

En nuestro motor de búsqueda considerando la consulta ejemplo la palabra “Mono” los resultados serían como se muestra en la siguiente figura.

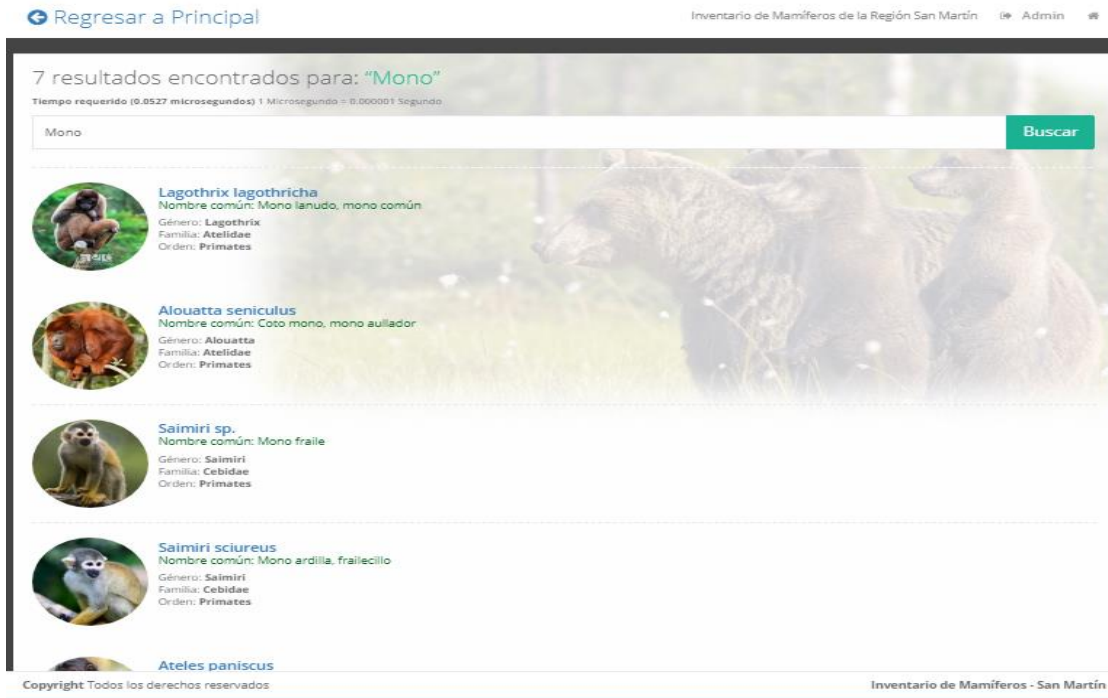
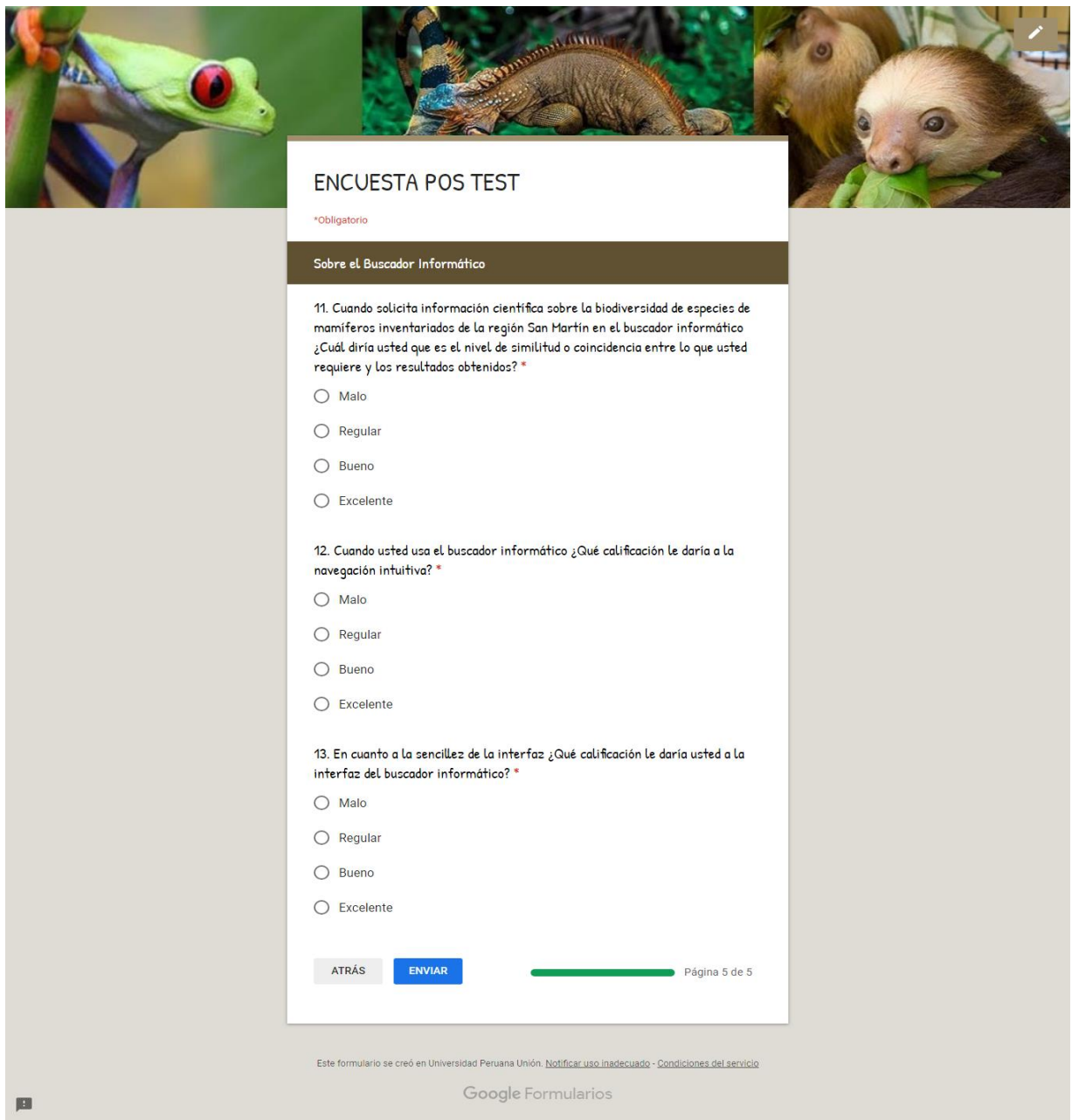


Figura 30 Resultados de la consulta "Mono"

#### 4.5. Evaluar postest

Finalmente, después de implementar el sistema para su uso respectivo en la difusión de la información, se aplicará un Post-Test para identificar la nueva percepción del usuario. Para la efectividad de la evaluación, utilizamos nuevamente formularios google.



The image shows a Google Form titled "ENCUESTA POS TEST" overlaid on a background of nature photos (a green frog, an iguana, and sloths). The form includes a header with the title and a red asterisk indicating mandatory questions. Below the title is a section header "Sobre el Buscador Informático". Three questions are listed, each with four radio button options: "Malo", "Regular", "Bueno", and "Excelente". The questions are:

- 11. Cuando solicita información científica sobre la biodiversidad de especies de mamíferos inventariados de la región San Martín en el buscador informático ¿Cuál diría usted que es el nivel de similitud o coincidencia entre lo que usted requiere y los resultados obtenidos? \*
- 12. Cuando usted usa el buscador informático ¿Qué calificación le daría a la navegación intuitiva? \*
- 13. En cuanto a la sencillez de la interfaz ¿Qué calificación le daría usted a la interfaz del buscador informático? \*

At the bottom of the form, there are two buttons: "ATRÁS" (grey) and "ENVIAR" (blue). A progress bar shows the current page as "Página 5 de 5". At the very bottom, there is a footer with the text "Este formulario se creó en Universidad Peruana Unión. [Notificar uso inadecuado](#) - [Condiciones del servicio](#)" and the "Google Formularios" logo.

Figura 31 Postest

## 4.6. Validar

En esta última se presenta el proceso de contrastación de la hipótesis.

### 4.6.1. Formulación de la hipótesis estadística:

**H<sub>0</sub>:**  $U_{pre} \geq U_{pos}$  significa que la puntuación calculada en el Pre Test es mayor o igual que la puntuación calculada del Pos Test.

**H<sub>a</sub>:**  $U_{pre} < U_{pos}$  significa que la puntuación calculada en el Pre Test es menor que la puntuación calculada del Pos Test.

Se pretende determinar que la puntuación calculada del Post Test sea mayor a la del Pre Test para de esta manera, demostrar que, con el uso del motor de búsqueda basado en el Modelo de Espacio Vectorial mejore la difusión de información científica de especies de mamíferos inventariados de la Región San Martín.

### 4.6.2. Nivel de significancia

Para la verificación de nuestra hipótesis utilizamos un nivel de significancia de  $\alpha=0.05$ .

### 4.6.3. Estadístico de prueba

Se utilizó  $n = 12$  porque el número de representantes de las instituciones que fueron evaluadas en el Pre Test como en el Pos Test son las mismas personas, por tanto, son muestras relacionadas y por ello el grado de libertad tomado en esta tesis será de  $n - 1 = 11$ .

$$U = \frac{Y_{pre} - Y_{pos}}{\sqrt{Sp^2 * \left(\frac{1}{n}\right)}}$$

*Donde:*

*Y<sub>pre</sub>: Media de tiempo del Pre Test*

*Y<sub>pos</sub>: Media de tiempo del Pos Test*

*Sp<sup>2</sup> : Varianza*

*n: número de encuestados*

Para calcular la Varianza usamos la siguiente fórmula:

$$Sp^2 = \frac{Spre^2 * n_{pre} + Spos^2 * n_{pos}}{n-1}$$

Donde:

$n_{pre}$ : Total de encuestados en el pre test

$n_{pos}$ : Total de encuestados en el post test

$Spre^2$ : Varianza del pre test

$Spos^2$ : Varianza del post test

Para calcular el “t” en la Tabla T-student usamos:

$$t_{(n-1; \infty)} = 1.7959$$

#### 4.6.4. Determinación de la región crítica:

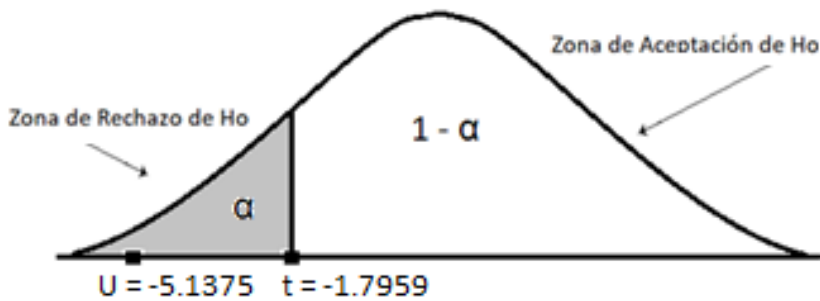


Figura 32 Región crítica

Para este caso la regla de decisión es la siguiente:

ACEPTAR  $H_0$  si  $U \in$  Zona de Aceptación

RECHAZAR  $H_0$  si  $U \in$  Zona de Rechazo

#### 4.6.5. Obtención del valor experimental:

Las siguientes tablas presentan la tabulación de los resultados obtenidos en el pre test y post test.

**4.6.5.1. Tabla de Pre-Test**

	Institución	Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4	Pregunta 5	Pregunta 6	Pregunta 7	Pregunta 8	Pregunta 9	Pregunta 10	PUNTAJE		
1	Administración Local del Agua - Tarapoto	3	2	2	3	3	3	3	2	2	2	25		
2	URKU Estudios Amazónicos	3	3	4	4	4	4	3	3	2	3	33		
3	Proyecto Especial Huallaga Central y Bajo Mayo	2	3	3	3	3	3	3	3	3	3	29		
4	Proyecto Especial Alto Mayo	3	2	3	3	3	3	3	2	2	2	26		
5	Instituto Nacional de Innovación Agraria	1	1	2	2	2	2	3	2	2	2	19		
6	CIMA Cordillera Azul	3	3	3	3	4	4	4	3	3	3	33		
7	EMAPA - San Martín	2	1	2	3	3	3	2	2	2	2	22		
8	SERNANP	3	2	3	4	4	4	4	4	3	4	35		
9	UGEL - SAN MARTÍN	3	2	3	2	2	3	3	2	3	2	25		
10	IIAP – SAN MARTÍN	2	2	2	2	2	2	2	2	2	2	20		
11	DRASAM	3	3	3	4	3	4	3	3	3	3	32		
12	SENASA	3	4	4	3	3	2	3	3	3	3	31		
												MEDIA	27.5	
													VARIANZA	29.54545455

4.6.5.2. *Tabla Post-Test*

#	Institución	Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4	Pregunta 5	Pregunta 6	Pregunta 7	Pregunta 8	Pregunta 9	Pregunta 10	PUNTAJE (Y)
1	Administración Local del Agua - Tarapoto	4	4	5	5	4	4	4	4	5	4	43
2	URKU Estudios Amazónicos	3	4	4	4	4	4	4	4	4	4	39
3	Proyecto Especial Huallaga Central y Bajo Mayo	5	5	3	3	3	3	3	3	3	3	34
4	Proyecto Especial Alto Mayo	3	3	2	3	3	3	3	2	2	3	27
5	Instituto Nacional de Innovación Agraria	3	3	3	3	3	3	3	4	5	3	33
6	CIMA Cordillera Azul	5	5	5	5	5	5	5	5	5	5	50
7	EMAPA - San Martín	4	5	5	4	4	4	4	4	5	4	43
8	SERNANP	3	4	4	4	4	4	4	4	4	4	39
9	UGEL - SAN MARTÍN	5	5	5	5	5	5	5	5	5	5	50
10	IIAP – SAN MARTÍN	5	3	5	3	5	3	5	3	5	3	40
11	DRASAM	5	5	5	3	5	2	5	5	5	5	45
12	SENASA	3	5	5	3	5	5	3	3	5	3	40
											MEDIA	40.25
											VARIANZA	45.29545455



#### 4.6.5.3. Fórmula del cálculo del Estadístico de Prueba (U):

$$U = \frac{Y_{pre} - Y_{pos}}{\sqrt{Sp^2 * (\frac{1}{n})}} = -5.1375$$

$$Sp^2 = \frac{Y_{pre} * n + Y_{pos} * n}{n - 1} = 73.9091$$

#### 4.6.5.4. Resumen de Datos Calculados:

Tabla 11

Resumen de datos calculados

Datos calculados		
1	varianza pre test:	29.5454545
2	varianza pos test:	45.2954545
3	Y <sub>pre</sub> (media pre test):	27.5
4	Y <sub>pos</sub> (media pos test):	40.25
5	n = n <sub>pre</sub> = n <sub>pos</sub> :	12
6	Sp <sup>2</sup> :	73.9090909
7	U (valor calculado):	-5.13750051
8	α (nivel de significancia):	0.05
9	T (11,0.05) (valor de tabla):	-1.7959

#### 4.6.6. Decisión

Entonces, siendo que  $U < t$  numéricamente **-5.1375 < -1.7959**

Por tanto, se Rechaza la Ho y se ACEPTA la Ha

#### 4.6.7. Conclusión

Entonces, se puede afirmar que el motor de búsqueda basado en el modelo de espacio vectorial mejora la difusión de información científica de mamíferos inventariados en la región San Martín.

## Capítulo 5 RESULTADOS Y DISCUSIÓN

En este capítulo se presenta los resultados obtenidos de las encuestas aplicadas a las 12 personas responsables de las diferentes instituciones dedicadas a la conservación de la biodiversidad. Para la obtención de estos datos recurrimos en algunos casos al apoyo de herramientas colaborativas de Google. A continuación, se detalla el resultado comparativo del Pre-Test y Post Test, dividido en sus tres dimensiones: la difusión bajo demanda, difusión documentaria, y medio de difusión utilizado, la cual es la que abordamos en esta investigación.

Para realizar el análisis de forma más sencilla, creamos un archivo en Microsoft Excel, en donde hicimos el vaciado de todos los datos obtenidos, para luego ser analizados por medio de tablas y gráficos.

### 5.1. Difusión bajo demanda

Las siguientes tablas 12 y 13 presentan un resumen de los resultados obtenidos de las encuestas antes de implementar el motor de búsqueda (pretest), y después de implementar el motor de búsqueda (postest). Los datos están agrupados bajo el marco de la dimensión denominada “Difusión bajo demanda”.

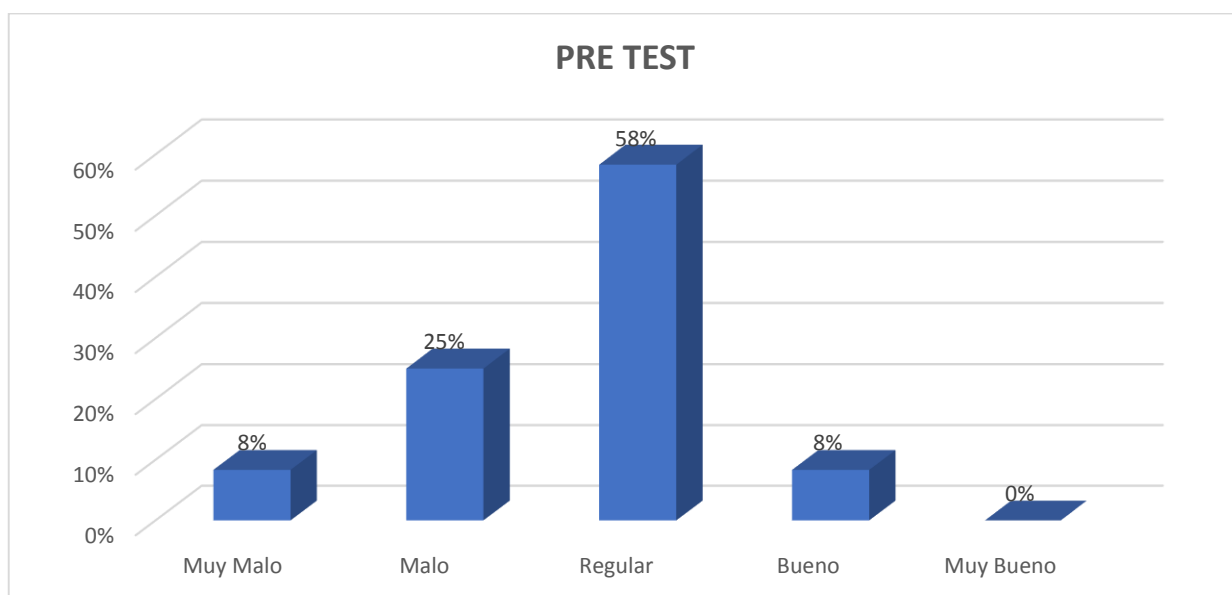
Tabla 12  
*Pretest difusión bajo demanda*

<b>Escala</b>	<b>Difusión bajo demanda</b>	<b>#</b>
Muy Malo	8%	1
Malo	25%	3
Regular	58%	7
Bueno	8%	1
Muy Bueno	0%	0
	100%	12

Tabla 13  
*Postest difusión bajo demanda*

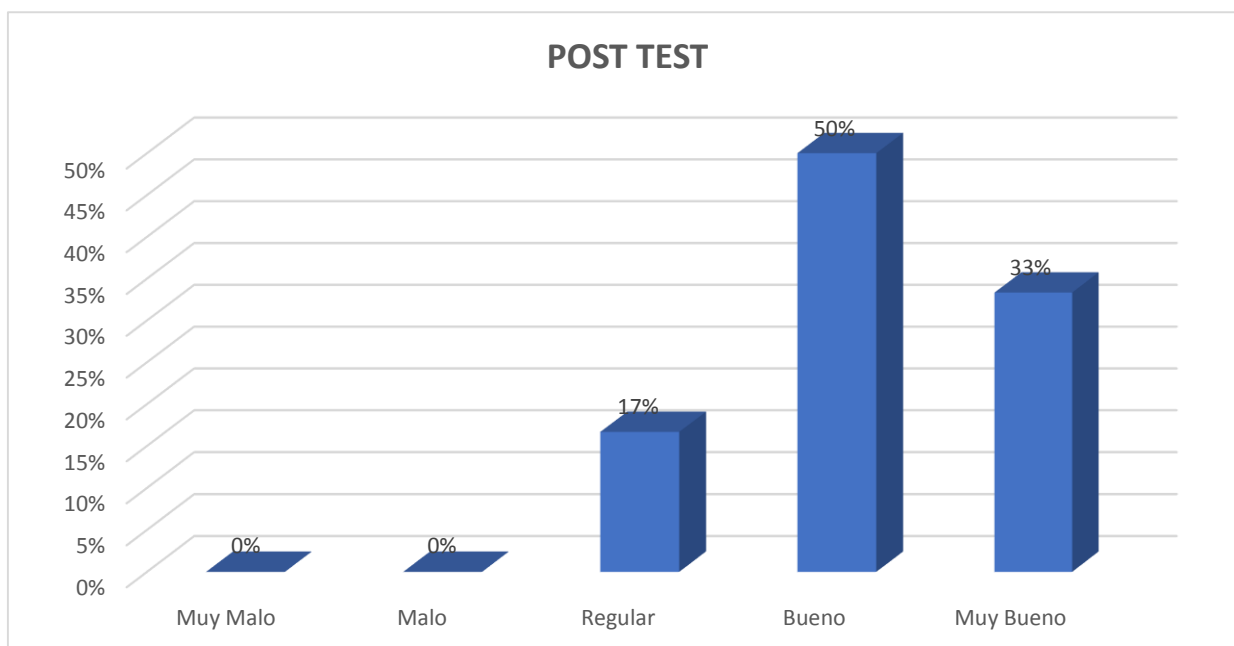
Escala	Difusión bajo demanda	#
Muy Malo	0%	0
Malo	0%	0
Regular	17%	2
Bueno	50%	6
Muy Bueno	33%	4
	100%	12

En el *Gráfico 1* se muestra que entre la población encuestada se encontró que el 33% califica a la difusión bajo demanda como mala, y otro 58% la considera regular. Esto indica que el 91% de la población considera a la difusión bajo demanda deficiente.



*Gráfico 1* Pretest difusión bajo demanda  
 Fuente: Elaboración propia

Sin embargo el Gráfico 2 muestra que del total de la población encuestada se encontró que el 30% califica a la difusión bajo demanda como muy buena, y otro 50% la considera buena. Esto indica que el 88% de la población que usa el motor de búsqueda, califica a la difusión bajo demanda como eficiente a la hora de realizar búsquedas, para obtener información científica.



*Gráfico 2 Postest difusión bajo demanda  
Fuente: Elaboración propia*

## **5.2. Difusión documentaria**

En las tablas 14 y 15 se presenta el resumen de los resultados obtenidos antes de implementar el motor de búsqueda (pretest), y después de implementar el motor de búsqueda (postest). Los datos están agrupados bajo el marco de la dimensión denominada “Difusión documentaria”.

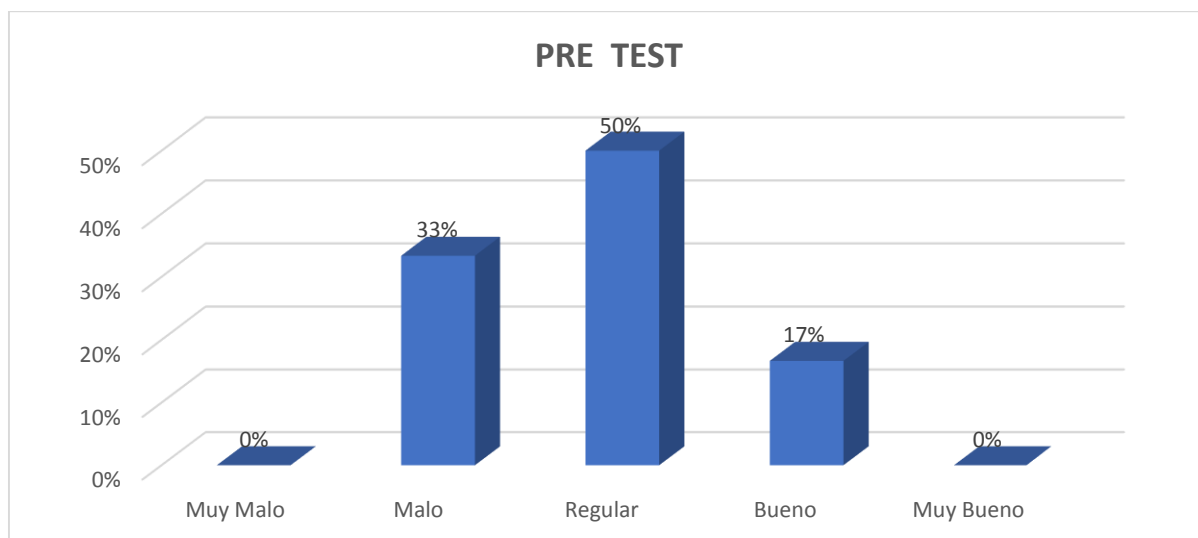
Tabla 14  
*Pretest difusión documentaria*

Escala	Difusión documentaria	#
Muy Malo	0%	0
Malo	33%	4
Regular	50%	6
Bueno	17%	2
Muy Bueno	0%	0
	100%	12

Tabla 15  
*Postest difusión documentaria*

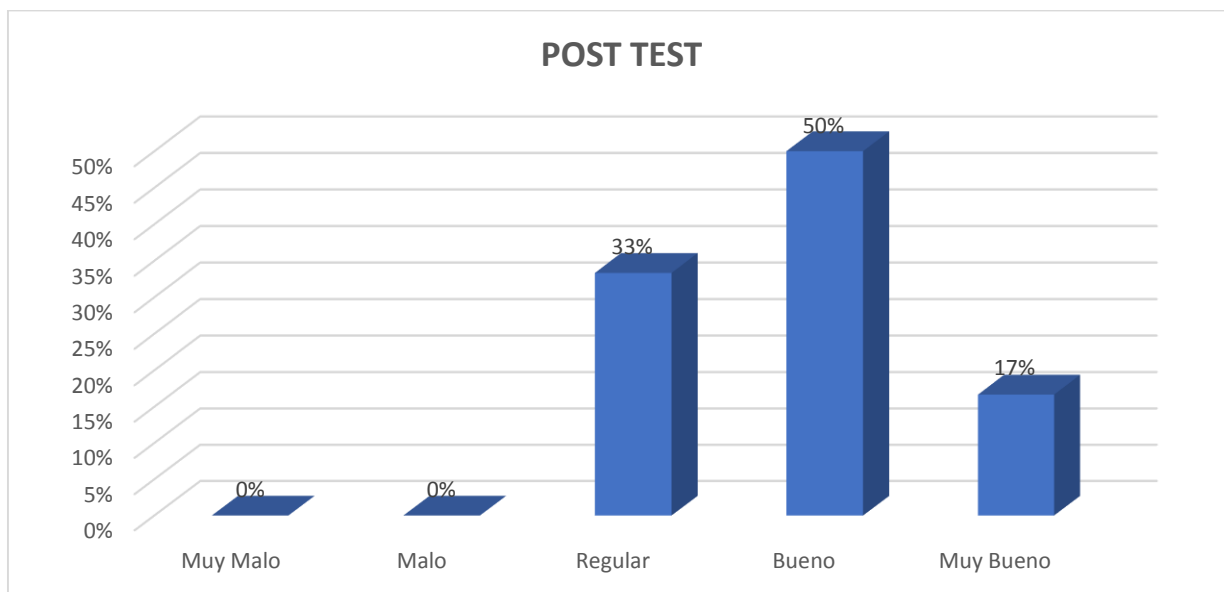
Escala	Difusión documentaria	#
Muy Malo	0%	0
Malo	0%	0
Regular	33%	4
Bueno	50%	6
Muy Bueno	17%	2
	100%	12

En el *Gráfico 3* se muestra que entre la población encuestada se encontró que el 33% califica a la difusión bajo documentaria como mala, y otro 50% la considera regular. Esto indica que el 83% de la población considera a la difusión documentaria deficiente.



*Gráfico 3 Pretest difusión documentaria*  
 Fuente: *Elaboración propia*

Sin embargo el *Gráfico 4* muestra que del total de la población encuestada el 17% califica a la difusión documentaria como muy buena, y otro 50% la considera buena. Esto indica que el 67% de la población que usa el motor de búsqueda, califica a la difusión documentaria como eficiente a la hora de realizar búsquedas, para obtener información científica.



*Gráfico 4 Postest difusión documentaria*  
Fuente: Elaboración propia

### 5.3. Medio de difusión utilizado

En las tablas 16 y 17 se presenta el resumen de los resultados obtenidos antes de implementar el motor de búsqueda (pretest), y después de implementar el motor de búsqueda (postest). Los datos están agrupados bajo el marco de la dimensión denominada “Medio de difusión utilizado”.

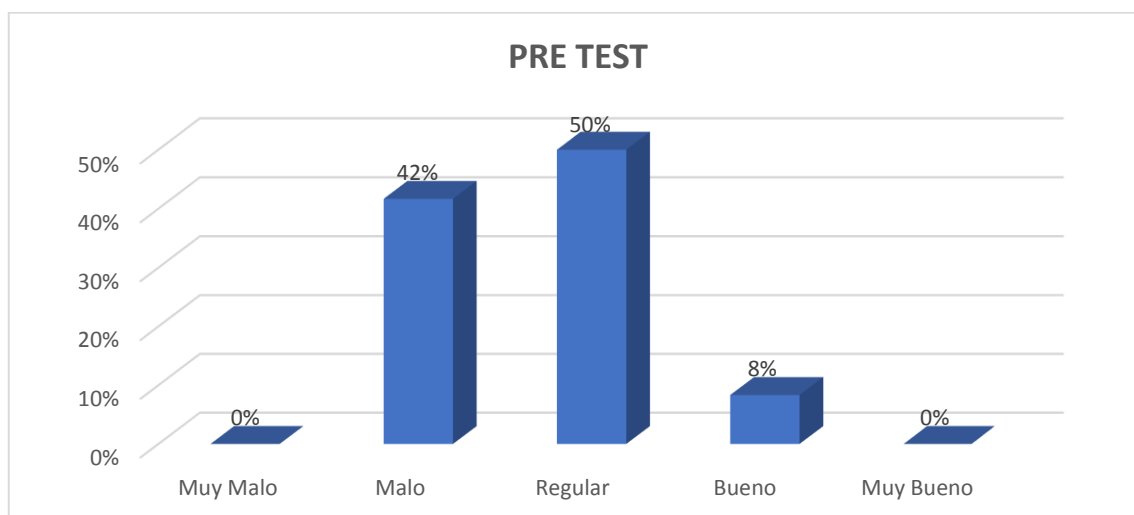
Tabla 16  
*Pretest medio de difusión utilizado*

Escala	Medio de difusión	#
Muy Malo	0%	0
Malo	42%	5
Regular	50%	6
Bueno	8%	1
Muy Bueno	0%	0
	100%	12

Tabla 17  
*Postest medio de difusión utilizado*

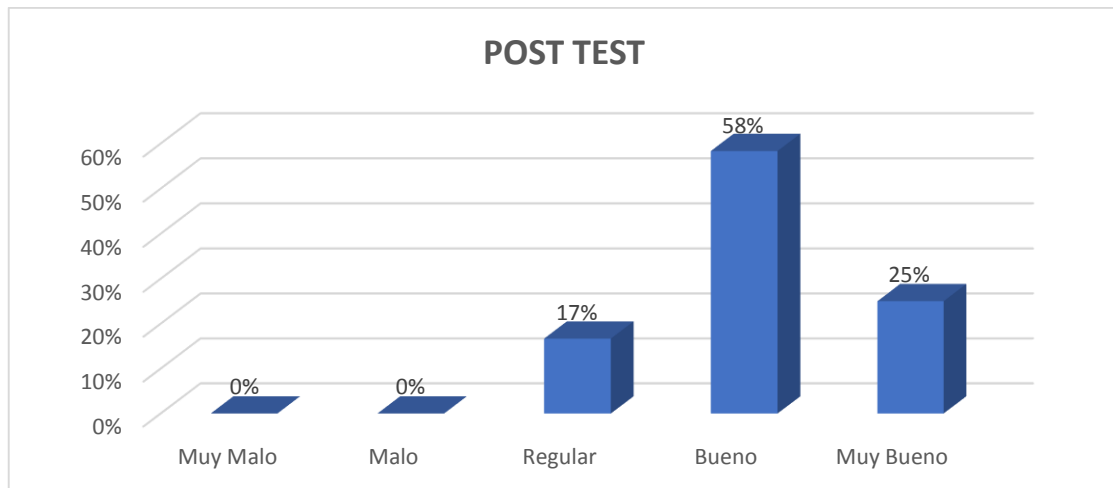
Escala	Medio de difusión	#
Muy Malo	0%	0
Malo	0%	0
Regular	17%	2
Bueno	58%	7
Muy Bueno	25%	3
	100%	12

En el *Gráfico 5* se muestra que entre la población encuestada se encontró que el 42% califica que el medio de difusión utilizado es malo, y otro 50% la considera regular. Esto indica que el 92% de la población considera que el medio de difusión es deficiente.



*Gráfico 5 Pretest medio de difusión utilizado*  
*Fuente: Elaboración propia*

Por otro lado, el *Gráfico 6* muestra que del total de la población encuestada, el 25% califica que el medio de difusión es muy bueno, y otro 58% la considera bueno. Esto indica que el 83% de la población que usa el motor de búsqueda, califica el medio de difusión como eficiente a la hora de realizar búsquedas, para obtener información científica.



*Gráfico 6 Postest medio de difusión utilizado  
Fuente: Elaboración propia*

#### **5.4. Sobre el Buscador Informático**

En esta sección se presenta el resultado de la dimensión “Sobre el motor de búsqueda”, esta fue agregada en el post test con el fin de medir el nivel de satisfacción, y apreciación en cuanto a la calidad, funcionalidad, y efectividad del motor de búsqueda informático.

A continuación, se presentan tres preguntas que forman parte de esta dimensión, tales se denominaron como nivel de similitud, navegación intuitiva, interfaz del buscador.

##### **5.4.1. Nivel de similitud**

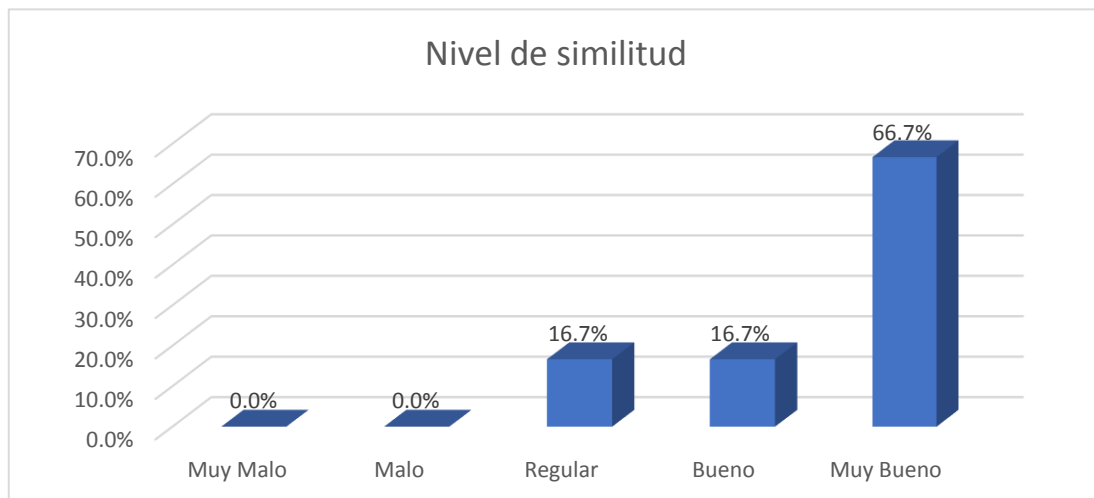
Cuando solicita información científica sobre la biodiversidad de especies de mamíferos inventariados de la región San Martín en el buscador informático ¿Cuál diría usted que es el nivel de similitud o coincidencia entre lo que usted requiere y los resultados obtenidos?



Tabla 18  
*Postest nivel de similitud*

Escala	Nivel de similitud	#
Muy Malo	0%	0
Malo	0%	0
Regular	16.7%	2
Bueno	16.7%	5
Muy Bueno	66.7%	5
	100%	12

El *Gráfico 7* muestra que, del total de la población encuestada, el 67% considera que el nivel de similitud es muy bueno, y otro 16.7% la considera bueno. Esto indica que el 83% de la población que usa el motor de búsqueda, califica el nivel de similitud entre la consulta y los resultados como bueno, dando fiabilidad a los resultados obtenidos.



*Gráfico 7 Postest nivel de similitud*  
*Fuente: Elaboración propia*

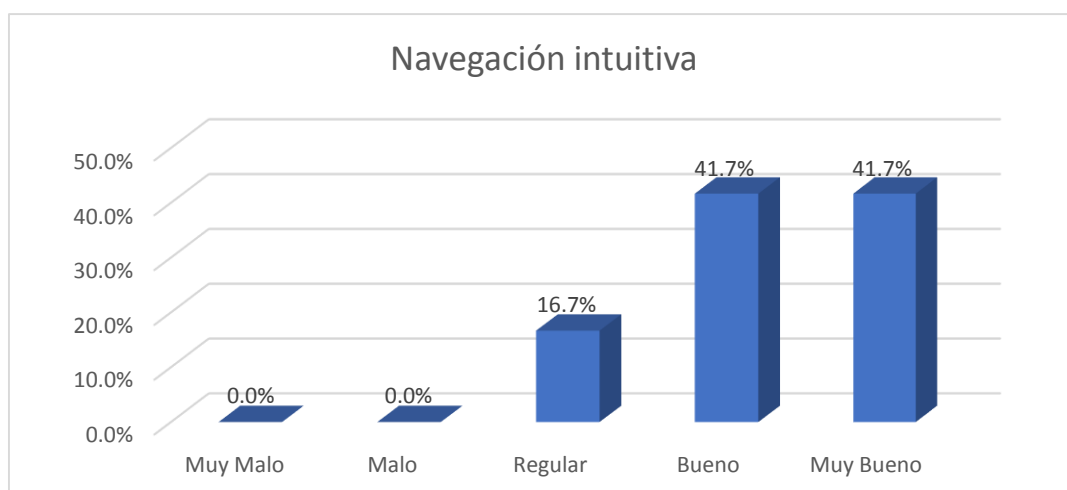
### 5.4.2. Navegación intuitiva

Pregunta 2: Cuando usted usa el buscador informático ¿Qué calificación le daría a la navegación intuitiva?

Tabla 19  
Postest navegación intuitiva

Escala	Navegación intuitiva	#
Muy Malo	0%	0
Malo	0%	0
Regular	16.7%	2
Bueno	41.7%	5
Muy Bueno	41.7%	5
	100%	12

El *Gráfico 8* muestra que, del total de la población encuestada, un 41.7% considera que la navegación intuitiva es muy buena, y otro 41.7% la considera buena. Esto indica que el 83% de la población que usa el motor de búsqueda, califica que la navegación es intuitiva, permitiendo inmediata obtención de información científica



*Gráfico 8 Postest navegación intuitiva*  
Fuente: Elaboración propia

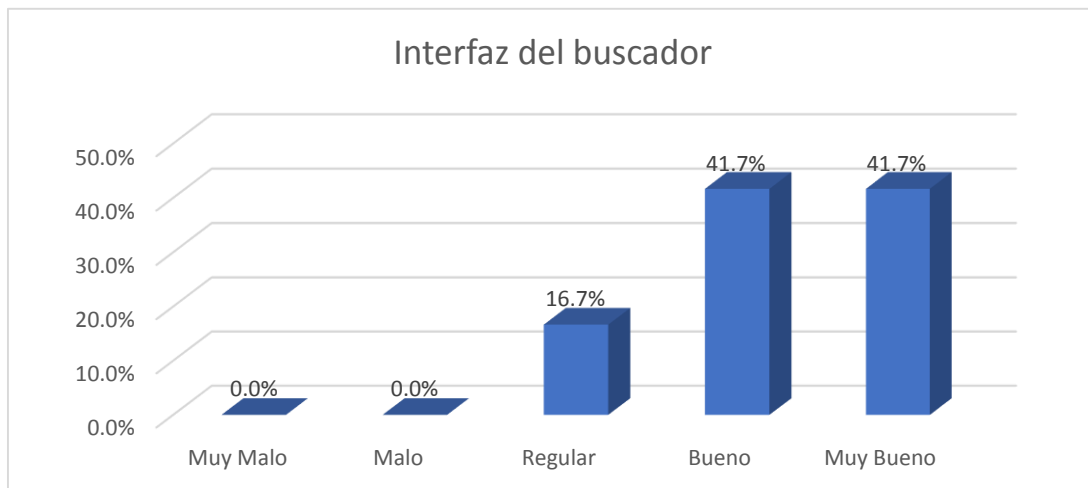
### 5.4.3. Interfaz del buscador

En cuanto a la sencillez de la interfaz ¿Qué calificación le daría usted a la interfaz del buscador informático?

Tabla 20  
Postest interfaz del navegador

Escala	Interfaz del navegador	#
Muy Malo	0%	0
Malo	0%	0
Regular	16.7%	2
Bueno	41.7%	5
Muy Bueno	41.7%	5
	100%	12

El *Gráfico 9* muestra que, del total de la población encuestada, un 41.7% califica que la interfaz del buscador es muy buena, y otro 41.7% la considera buena. Esto indica que el 83% de la población que usa el motor de búsqueda, considera que la interfaz del buscador es ágil y sencilla, permitiendo una comunicación eficaz entre el sistema y los usuarios.



*Gráfico 9 Postest interfaz del navegador*  
Fuente: Elaboración propia

### 5.5. Influencia de la variable X sobre la variable Y

Utilizamos el  $R^2$  que es el Coeficiente de Determinación, que se encuentra entre los valores de 0 a 1, el cual indica:

- Si el valor de  $R^2$  está más cercano a 0 (cero) entonces la influencia de la variable X sobre Y es débil.
- Si el valor de  $R^2$  está más cercano a 1 (uno) entonces la influencia de la variable X sobre Y es fuerte.

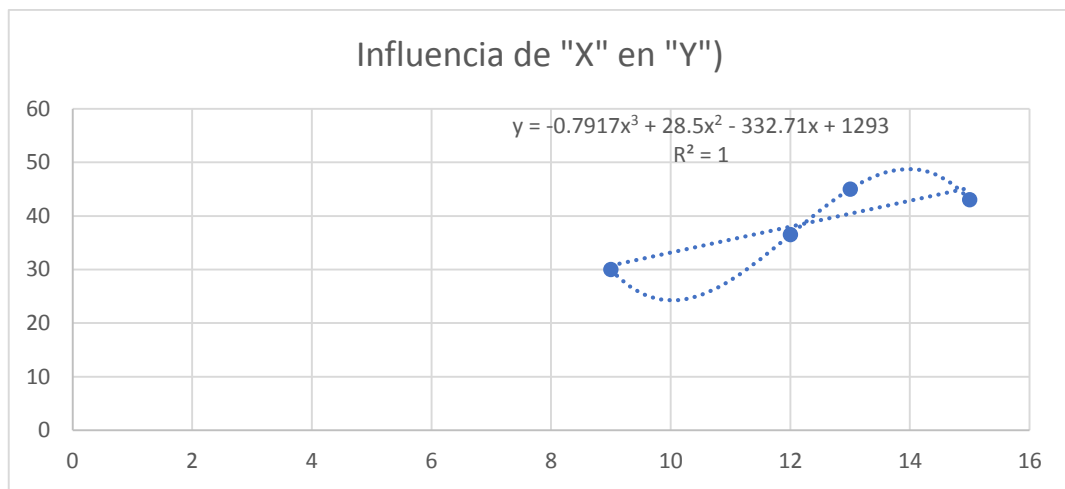


Gráfico 10 Influencia de x en y  
Fuente: Elaboración propia

Concluimos entonces que el valor de  $R^2 = 1$  es el coeficiente de determinación y se define como la proporción de la varianza total de la variable explicada por la regresión. En este trabajo se ha obtenido el valor "1" que indica que existe una relación perfecta entre las variables, por lo tanto, existe una fuerte relación entre ellas.

## **Capítulo 6**

### **CONCLUSIONES Y RECOMENDACIONES**

#### **6.1. Conclusiones**

- Del total de representantes encuestados en el pre test, referente al medio de difusión utilizado, el 42% lo califica como malo, y un 50% lo califica como regular, por lo cual concluimos que el medio de difusión es regular; en cuanto al post test se encontró que, un 25% de los representantes califica el medio de difusión como muy bueno, y un 58% lo califica como bueno, esto confirma que el uso del motor de búsqueda como medio de difusión para la información científica de mamíferos, logró una excelente aprobación.
- Referente al nivel de percepción del motor de búsqueda, los resultados muestran que, existe un 84% de representantes de las instituciones que manifiesta su aprobación en cuanto al uso del motor de búsqueda basado en el Modelo de Espacio Vectorial, y solo existe un 17% que lo considera regular.
- En cuanto a la relación del motor de búsqueda y sus parámetros, el total de encuestados opina que: la similitud de la consulta es muy buena con un 66.7%, su navegación es intuitiva es muy buena con un 41.7, y la sencillez de interfaz es muy buena con 41.7%.
- El modelo de espacio vectorial en la recuperación de la información es muy importante ya que, permite que el acceso a grandes volúmenes de información sea tan sencillo al momento de realizar una consulta, pues utilizando términos claves obtendremos información relevante y sobre todo de nuestro interés como el caso de la información científica de mamíferos.
- La implementación de dichos sistemas de RI nos desafía a sumar esfuerzos para desarrollar en el futuro sistemas de búsqueda de información más eficientes para gestionar información más confiable y de calidad.

## **6.2. Recomendaciones**

- El registro de la información futura, sea confiable y validada por taxónomos especialistas en el área de investigación.
- Desarrollar modelos computacionales que aporten al desarrollo sostenible de la región San Martín.
- Compartir el acceso al motor de búsqueda con la comunidad científica del Perú.
- Culturizar el uso de nuevas tecnologías de información en el campo de investigación.

## Capítulo 7 REFERENCIAS BIBLIOGRAFICAS

- Baeza Yates, R., & Ribeiro Neto, B. (1999). Modern information retrieval. *New York*, 9, 513. <https://doi.org/10.1080/14735789709366603>
- Barragán Reyes, A. M. (2008). *Un Método para Recuperación de Información en Documentos Orales basado en Codificación Fonética* (Instituto Nacional de Astrofísica, Óptica y Electrónica). Retrieved from <https://inaoe.repositorioinstitucional.mx/jspui/bitstream/1009/555/1/ReyesBMA.pdf>
- Blázquez Ochando, M. (2012). Técnicas avanzadas de recuperación de información: Frecuencias y pesos de los términos de un documento. Retrieved January 10, 2019, from <http://ccdoc-tecnicasrecuperacioninformacion.blogspot.com/2012/11/frecuencias-y-pesos-de-los-terminos-de.html>
- Blázquez Ochando, M. (2013). Técnicas avanzadas de recuperación de información. In *Técnicas avanzadas de recuperación de información: procesos, técnicas y métodos*. Retrieved from <http://mblazquez.es/libros-y-manuales-de-la-documentacion-no-4-tecnicas-avanzadas-de-recuperacion-de-informacion/>
- Bordignon, F., & Tolosa Chacón, G. (2007). RECUPERACIÓN DE INFORMACIÓN: UN ÁREA DE INVESTIGACIÓN EN CRECIMIENTO. *Red de Revistas Científicas de América Latina y El Caribe, España y Portuga*. Retrieved from <https://www.redalyc.org/articulo.oa?id=181414865002>
- Castillo Blasco, L. (2014). Temario de Documentación. Universidad de Valencia - Bibliopos: Biblioteca de recursos para Bibliotecarios y Opositores. Retrieved January 10, 2019, from <https://www.bibliopos.es/temario-de-documentacion-universidad-de-valencia/>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. R., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. Retrieved from <https://www.semanticscholar.org/paper/CRISP-DM-1.0%3A-Step-by-step-data-mining-guide-Chapman-Clinton/54bad20bbc7938991bf34f86dde0babfbd2d5a72>

- Dorado Nájera, A. (2010). *¿Que es la Biodiversidad?* (Fundación). Retrieved from <http://www.ecomilenio.es/wp-content/uploads/2010/10/que-es-la-biodiversidad-web.pdf>
- Galán Cortina, V. (2015). *Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno universitario* (Universidad Carlos III de Madrid). Retrieved from <https://e-archivo.uc3m.es/handle/10016/22198>
- Hernández, R., Fernández, C., & Baptista, P. (2014). Metodología de la investigación. In *Journal of Chemical Information and Modeling* (Vol. 53). <https://doi.org/10.1017/CBO9781107415324.004>
- Huacani mamani, O. R. (2014). *Recuperación de la información empleando el modelo de espacio vectorial en la Gestión Documentaria para la Unidad de Resoluciones de la Universidad Nacional del Altiplano - Puno* (Universidad Nacional del Altiplano). Retrieved from <http://renati.sunedu.gob.pe/handle/sunedu/131579>
- Jiménez Sierra, C., Torres Orozco, R., & Corcuera Martínez del Río, P. (2010). *Biodiversidad: Una alerta*. 3(36), 9–16. Retrieved from [http://www.uam.mx/difusion/casadeltiempo/36\\_iv\\_oct\\_2010/casa\\_del\\_tiempo\\_eIV\\_num36\\_09\\_16.pdf](http://www.uam.mx/difusion/casadeltiempo/36_iv_oct_2010/casa_del_tiempo_eIV_num36_09_16.pdf)
- KDnuggets. (2014). CRISP-DM, still the top methodology for analytics, data mining, or data science projects. Retrieved March 24, 2019, from <https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>
- Luque Rodriguez, M. (2005). Modelos de Recuperación de la Información basados en Información Lingüística Difusa y Algoritmos Evolutivos. Mejorando la Representación de las Necesidades de Información. (Universidad de Granada). Retrieved from <https://dialnet.unirioja.es/servlet/tesis?codigo=19029>
- Martínez Comeche, J. A. (2006). *Los modelos clásicos de Recuperación de información y su vigencia*. (2005), 1–19. Retrieved from <http://eprints.ucm.es/5979/>
- Martínez Méndez, F. J. (2002). Propuesta y desarrollo de un modelo para la evaluación de la recuperación de información en internet. Retrieved from




<http://www.tdx.cat/handle/10803/10904>

- Merlino santesteban, C. (2003). *Análisis de conectividad en la recuperación de información web*. 113–119. Retrieved from [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652003000300013&lng=en&nrm=iso&tlng=es](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652003000300013&lng=en&nrm=iso&tlng=es)
- Moine, J. M., Haedo, A. S., & Gordillo, S. (2011). *Estudio comparativo de metodologías para minería de datos*. 1–4. Retrieved from <http://sedici.unlp.edu.ar/handle/10915/20034>
- Rodríguez Camiño, R. (2003). Motores de búsqueda sobre salud en Internet. *Acimed, Scielo*, 1–21. Retrieved from [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1024-94352003000500002](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352003000500002)
- Romero, L. (2005). Divulgación científica y la biodiversidad. *Revista Peruana de Biología, Scielo Perú*, 12(2), 9933. Retrieved from [http://www.scielo.org.pe/scielo.php?script=sci\\_arttext&pid=S1727-99332005000200002](http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1727-99332005000200002)
- Salinas, P. J. (2012). *Metodología de la investigación*. Retrieved from <https://download.e-bookshelf.de/download/0003/5873/05/L-G-0003587305-0006913492.pdf>
- SERFOR. (2016). *Primer informe parcial del Inventario Nacional Forestal y de Fauna Silvestre*. Retrieved from <https://www.serfor.gob.pe/conoce/informe-del-primer-inventario-nacional-forestal-y-de-fauna-silvestre>
- StatCounter Global Stats. (2019). Search Engine Market Share Worldwide. Retrieved April 7, 2019, from <http://gs.statcounter.com/search-engine-market-share#monthly-201803-201903-bar>
- Tacillo Yauli, E. F. (2016). *Metodología de la Investigación Científica*. Retrieved from [https://alicia.concytec.gob.pe/vufind/Record/UJBM\\_5c70c87bd5198f530e5c6079d5061ea7](https://alicia.concytec.gob.pe/vufind/Record/UJBM_5c70c87bd5198f530e5c6079d5061ea7)
- Zambrano Guerra, V. (2009). *Diseño y Producción de sitio Web dedicado a la*

*fotografía documental Ecológica. Primera Galería: Cañaote: “Una segunda oportunidad de vivir”* (Universidad Central de Venezuel). Retrieved from <http://hdl.handle.net/123456789/2778>

## Capítulo 8 ANEXOS

### 8.1. Pretest



**Universidad Peruana Unión**  
**Proyecto de Investigación**  
**“Motor de Búsqueda basado en el Modelo de Espacio Vectorial para la Difusión de la Información Científica de las Especies de Mamíferos Inventariados en la Región San Martín.”**

**ENCUESTA PRE TEST**

INSTITUCIÓN: .....

FECHA: ...../...../.....

**Instrucciones:**  
Estimado encuestado, solicitamos tu colaboración para el desarrollo del presente cuestionario. Cada pregunta tiene 5 alternativas de respuesta.  
Marca con una equis (X) la letra que corresponda, según su criterio, a la opción que represente más su opinión:

**SOBRE LA DIFUSIÓN DE INFORMACIÓN**

**DIFUSIÓN BAJO DEMANDA**

1. Considera que, en la región San Martín, la organización de la información científica de la biodiversidad de especies de mamíferos inventariados para su consulta es:

a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

2. La posibilidad de obtener una copia fiel de información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín es:

a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

3. La posibilidad de buscar información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín utilizando diferentes criterios (como nombre común, nombre científico, palabras clave, etc), lo considera:

a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

*Anexos I Pre Test - P1*

#### DIFUSIÓN DOCUMENTARIA

4. La información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín que ofertan otras instituciones que manejan este tipo de información, lo considera:

- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

5. La modalidad (revistas, conferencias, cursos, videos informativos, etc) que se usa para transmitir la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín, lo considera:

- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

6. Cuando busca una información en particular, la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín sugeridos por otras instituciones que manejan este tipo de información, lo considera:

- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

#### MEDIO DE DIFUSIÓN UTILIZADO

7. El libre acceso a la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín, lo considera:

- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

8. La posibilidad de obtener toda la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín, lo considera:

- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

9. El tiempo de respuesta en obtener la información requerida, lo considera:


- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

10. La disponibilidad de la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín, lo considera:

- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

MUCHAS GRACIAS.

## 8.2. Postest



**Universidad Peruana Unión – Tarapoto**  
**Proyecto de Investigación**  
**“Motor de Búsqueda basado en el Modelo de Espacio Vectorial para la Difusión de la Información Científica de las Especies de Mamíferos Inventariados en la Región San Martín”**

**ENCUESTA POS TEST**

INSTITUCIÓN: .....

FECHA:...../...../.....

**Instrucciones:**  
Estimado encuestado, solicitamos tu colaboración para el desarrollo del presente cuestionario. Cada pregunta tiene 5 alternativas de respuesta.  
Marca con una equis (X) la letra que corresponda, según su criterio, a la opción que represente más su opinión:

**SOBRE LA DIFUSIÓN DE INFORMACIÓN**

**DIFUSIÓN BAJO DEMANDA**

1. Considera que, en la región San Martín, la organización de la información científica de la biodiversidad de especies de mamíferos inventariados para su consulta es:  
a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno
  
2. La posibilidad de obtener una copia fiel de información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín es:  
a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno
  
3. La posibilidad de buscar información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín utilizando diferentes criterios (como nombre común, nombre científico, palabras clave, etc), lo considera:  
a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

*Anexos 3 Post Test - P1*

#### DIFUSIÓN DOCUMENTARIA

4. La información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín que ofertan otras instituciones que manejan este tipo de información, lo considera:

- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

5. La modalidad (Motor de búsqueda) que se usa para transmitir la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín, lo considera:

- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

6. Cuando busca una información en particular, la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín sugeridos por otras instituciones que manejan este tipo de información, lo considera:

- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

#### MEDIO DE DIFUSIÓN UTILIZADO

7. El libre acceso a la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín, lo considera:

- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

8. La posibilidad de obtener toda la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín, lo considera:

- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

9. El tiempo de respuesta en obtener la información requerida, lo considera:

- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

10. La disponibilidad de la información científica de la biodiversidad de especies de mamíferos inventariados de la región San Martín, lo considera:

- a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

### **SOBRE EL BUSCADOR INFORMÁTICO**

1. Cuando solicita información científica sobre la biodiversidad de especies de mamíferos inventariados de la región San Martín en el motor de búsqueda ¿Cuál diría usted que es el nivel de similitud o coincidencia entre lo que usted requiere y los resultados obtenidos?  
a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno
  
2. Cuando usted usa el buscador informático ¿Qué calificación le daría a la navegación intuitiva?  
a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno
  
3. En cuanto a la sencillez de la interfaz ¿Qué calificación le daría usted a la interfaz del buscador informático?  
a) Muy Malo      b) Malo      c) Regular      d) Bueno      e) Muy bueno

**MUCHAS GRACIAS.**

### 8.3. Preguntas de entrevista

PREGUNTAS DE ENTREVISTA		
1	¿Cómo se ha obtenido el inventario de fauna silvestre de la Región San Martín?	
2	¿Quién es el responsable de hacer el inventario de fauna silvestre de la Región San Martín?	
3	¿Cuánto tiempo toma realizar la difusión?	
4	¿Cada cuánto tiempo se realiza una actualización del inventario de fauna silvestre de la Región San Martín?	
5	¿Cada cuánto tiempo se realiza una actualización del inventario de fauna silvestre de la Región San Martín?	
6	¿Cómo se realiza el proceso de difusión de información científica de fauna silvestre inventarios de la Región San Martín?	
7	¿Qué áreas u oficinas del IIAP están involucradas con la difusión de información científica de fauna silvestre inventariadas de la Región San Martín?	
8	¿Cuánto le cuesta al IIAP hacer la difusión?	
9	¿Qué dificultades se tiene en el proceso de difusión?	
10	¿Cuál es la población beneficiaria de estos datos?	
11	¿Considera que esa difusión es efectiva?, Por qué?	
12	¿Qué etapas de la difusión considera que se pueden mejorar?, Por qué?	

*Anexos 6 Preguntas de entrevistas*



#### 8.4. Validación de juicio de expertos

##### INSTRUMENTO PARA LA VALIDEZ DE CONTENIDO (JUICIO DE EXPERTOS)

El presente instrumento tiene como finalidad evaluar la conformidad del usuario respecto a la difusión de la información científica, el mismo será aplicado a representantes de Instituciones de la Región San Martín, quienes constituyen la muestra en estudio de la validación del instrumento titulado: INVENTARIO DE MAMÍFEROS

##### Instrucciones


La evaluación requiere de la lectura detallada y completa de cada uno de los ítems propuestos a fin de cotejarlos de manera cualitativa con los criterios propuestos relativos a: **relevancia o congruencia con el contenido, claridad en la redacción, tendenciosidad o sesgo en su formulación y dominio del contenido**. Para ello deberá asignar una valoración si el ítem presenta o no los criterios propuestos, y en caso necesario se ofrece un espacio para las observaciones hubiera.

Juez N°: 01 Fecha actual: 18/06/2019

Nombres y Apellidos de Juez John Clark Santa María Pinedo

Institución donde elabora: Universidad Nacional de San Martín

Años de experiencia profesional o científica: 11 años

  
Firma y Sello

Anexos 7 Validación del Juez n° 1

**INSTRUMENTO PARA LA VALIDEZ DE CONTENIDO  
(JUICIO DE EXPERTOS)**

El presente instrumento tiene como finalidad evaluar la conformidad del usuario respecto a la difusión de la información científica, el mismo será aplicado a representantes de Instituciones de la Región San Martín, quienes constituyen la muestra en estudio de la validación del instrumento titulado: INVENTARIO DE MAMÍFEROS

**Instrucciones**

La evaluación requiere de la lectura detallada y completa de cada uno de los ítems propuestos a fin de cotejarlos de manera cualitativa con los criterios propuestos relativos a: **relevancia o congruencia con el contenido, claridad en la redacción, tendenciosidad o sesgo en su formulación y dominio del contenido**. Para ello deberá asignar una valoración si el ítem presenta o no los criterios propuestos, y en caso necesario se ofrece un espacio para las observaciones hubiera.

Juez N°: 02 Fecha actual: 11/06/19

Nombres y Apellidos de Juez: Miguel Valles Coral

Institución donde elabora: Universidad Nacional de San Martín

Años de experiencia profesional o científica: 16 años

Miguel Ángel Valles Coral  
ING. QUÍMICO  
CIP. 5213

Firma y Sello

*Anexos 8 Validación del Juez n° 2*

**INSTRUMENTO PARA LA VALIDEZ DE CONTENIDO  
(JUICIO DE EXPERTOS)**

El presente instrumento tiene como finalidad evaluar la conformidad del usuario respecto a la difusión de la información científica, el mismo será aplicado a representantes de Instituciones de la Región San Martín, quienes constituyen la muestra en estudio de la validación del instrumento titulado: INVENTARIO DE MAMÍFEROS

**Instrucciones**

La evaluación requiere de la lectura detallada y completa de cada uno de los ítems propuestos a fin de cotejarlos de manera cualitativa con los criterios propuestos relativos a: **relevancia o congruencia con el contenido, claridad en la redacción, tendenciosidad o sesgo en su formulación y dominio del contenido**. Para ello deberá asignar una valoración si el ítem presenta o no los criterios propuestos, y en caso necesario se ofrece un espacio para las observaciones hubiera.

Juez N°: 03 Fecha actual: 03-06-19

Nombres y Apellidos de Juez Joyse Baldwin Huamán Labán

Institución donde elabora: Universidad Peruana Unión

Años de experiencia profesional o científica: 3 años


Firma y Sello

*Anexos 9 Validación del Juez n° 3*

## **8.5. Detalle del desarrollo del motor de búsqueda**

### **8.5.1. Funcionalidades del negocio**

- Registrar Usuario
- Registrar Mamíferos
- Buscar Mamíferos
- Generar lista Invertida

### **8.5.2. Modelado del negocio**

#### **8.5.2.1. Actores**

- Usuario
- Administrador

#### **8.5.2.2. Casos de uso**

- Consulta de mamíferos
- Mantenimiento
- Seguridad

### 8.5.2.3. Modelo de caso de uso de negocio

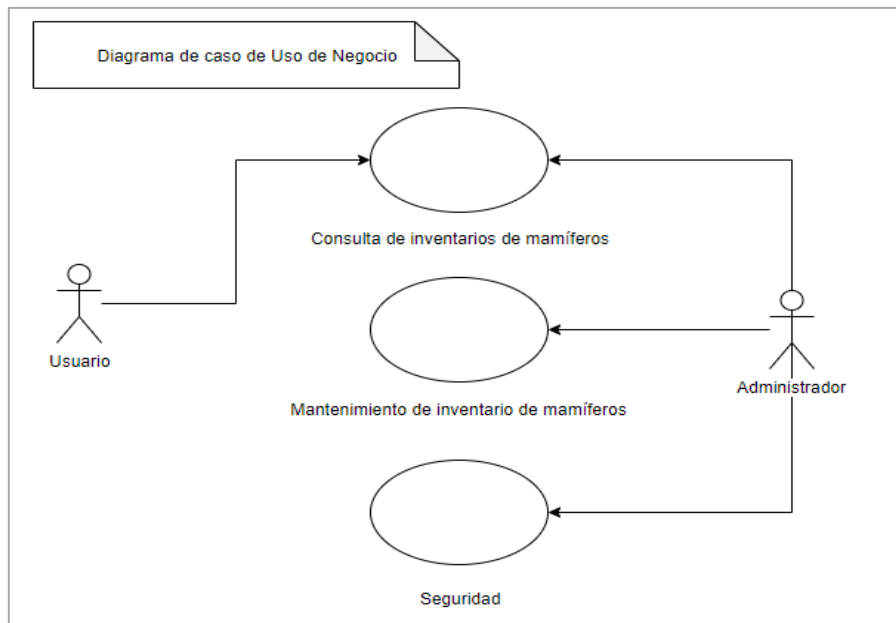


Figura 33 Diagrama de caso de uso de negocio.  
Fuente: Elaboración propia

Tabla 21  
Consulta de inventario de mamíferos

Caso de Uso del negocio	Consulta de inventarios de mamíferos
Actor	Cliente: Usuario del sistema
Resumen	Este caso de uso de negocio “Consulta de inventarios de mamíferos”, permite al actor tener un listado de las tesis que corresponden a la relevancia de los términos de su búsqueda basado en los campos “Nombre común, y Nombre científico”.
Propósito	Buscar inventarios de mamíferos
Objetivo	Obtener un listado de inventarios de mamíferos basado en los términos de búsqueda.

Tabla 22  
*Mantenimiento de inventarios de mamíferos*

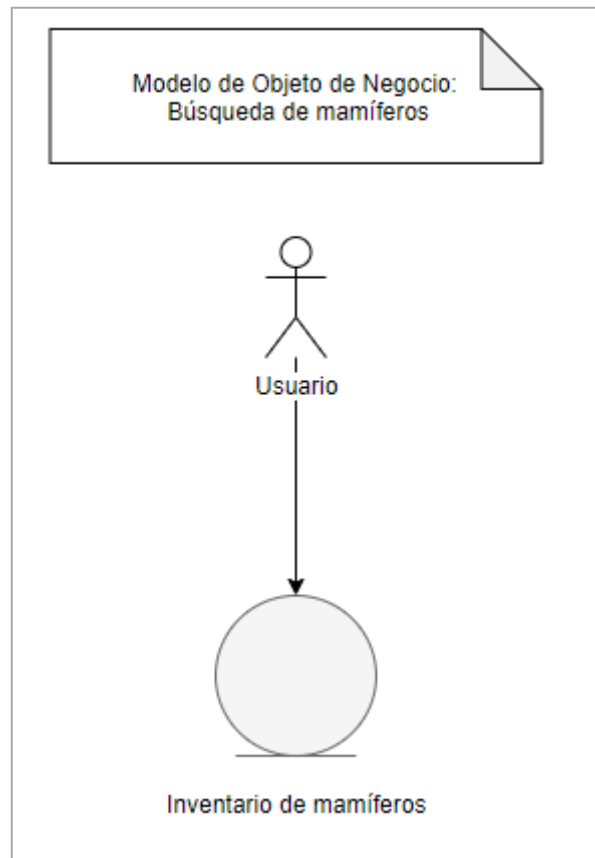
Caso de Uso del negocio	Mantenimiento de inventarios de mamíferos
Actor	Administrador
Resumen	Este caso de uso de Negocio “Mantenimiento”, permite el total control sobre los datos de los inventarios de mamíferos. Estos datos permiten que el usuario pueda generar una Lista invertida válida que sirva de índice para llevar a cabo la consulta de registros de tesis.
Propósito	Registrar inventarios de mamíferos Generar lista invertidas
Objetivo	Tener control sobre los inventarios de mamíferos

Tabla 23  
*Seguridad*

Caso de Uso del negocio	Seguridad
Actor	Administrador
Resumen	Este caso de uso de Negocio “Seguridad”, permite el total control sobre los usuarios que manipulan los datos de mamíferos en el sistema, dichos usuarios tienen acceso a registrar, editar, y eliminar los datos de los mamíferos.
Propósito	Registrar de usuario
Objetivo	Realizar el control de los accesos de Usuarios

**8.5.2.4. Modelo de objeto de negocio (MON)**

**8.5.2.4.1. MON Búsqueda de inventarios**



*Figura 34 MON Búsqueda de inventarios.  
Fuente: Elaboración propia*

#### 8.5.2.4.2. MON Mantenimiento

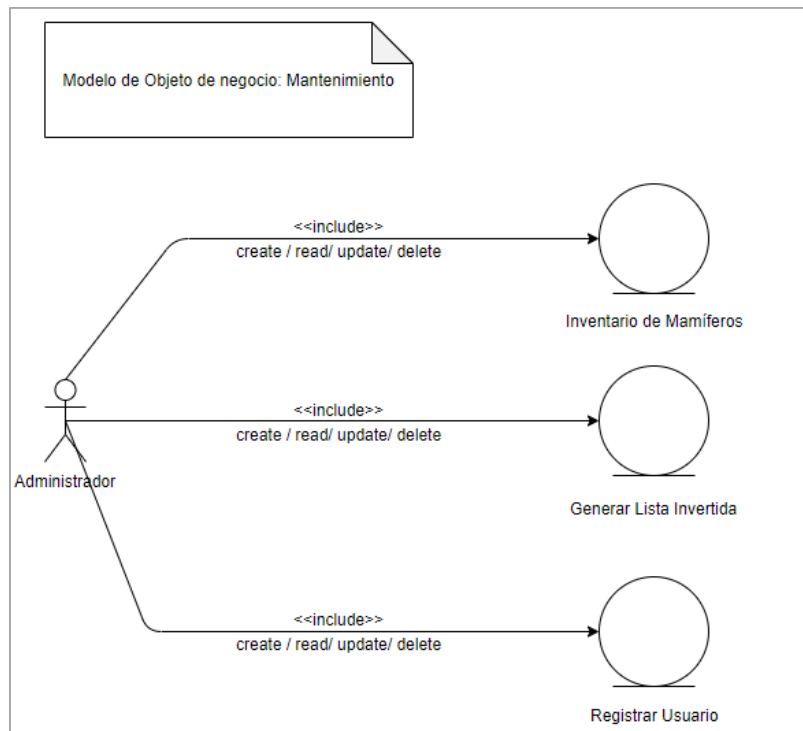


Figura 35 MON Mantenimiento  
Fuente: Elaboración propia

#### 8.5.2.4.3. MON Seguridad

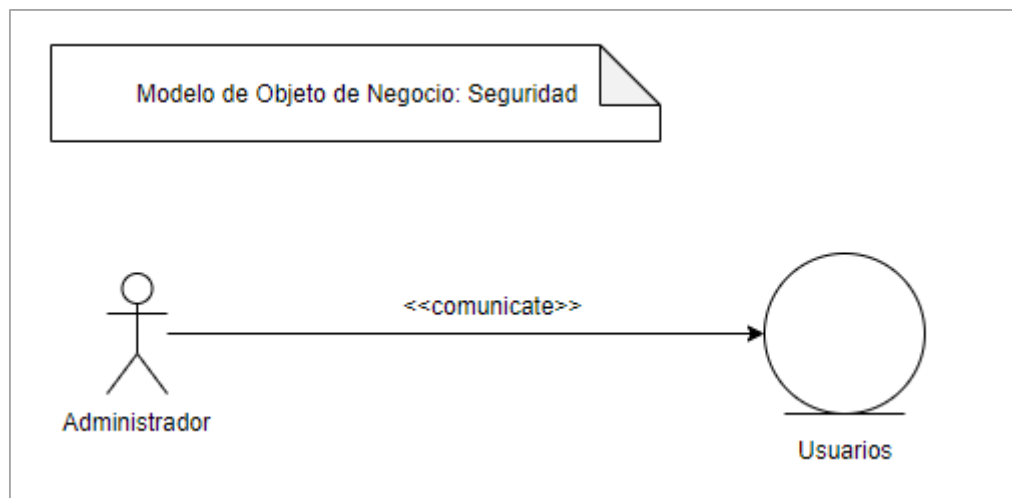


Figura 36 MON Seguridad.  
Fuente: Elaboración propia



### 8.5.2.5. Modelos de caso de Uso de requerimiento

#### 8.5.2.5.1. Modelo de casos de uso de requerimiento (MCUR)

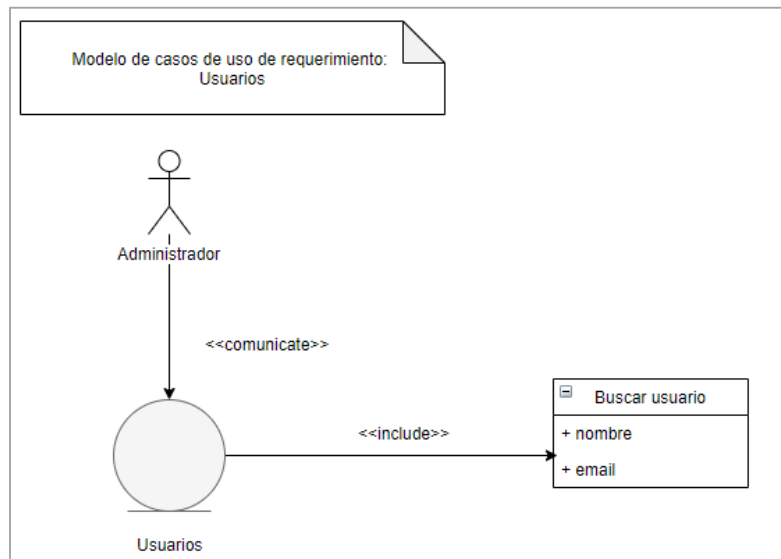


Figura 37 MCUR Seguridad.  
Fuente: Elaboración propia

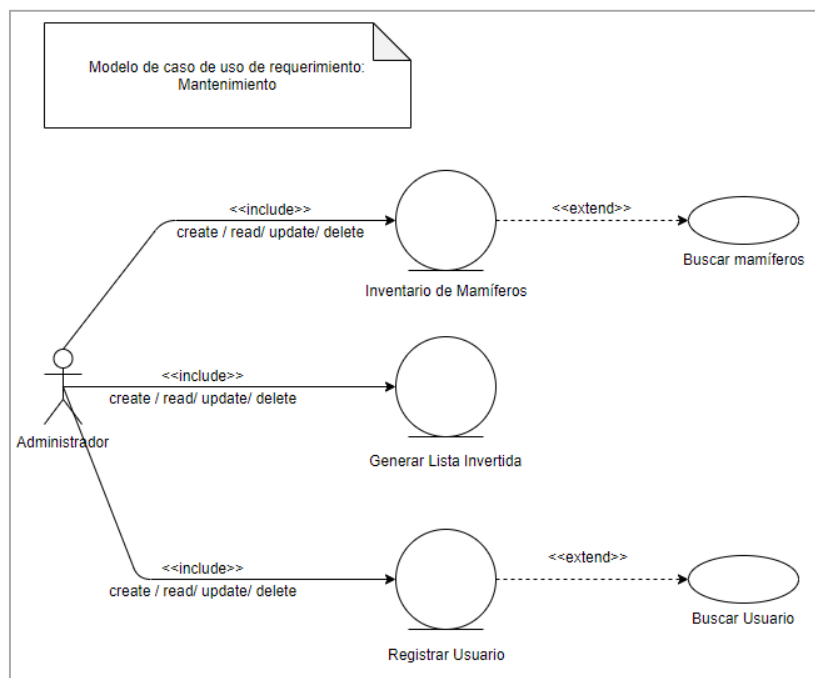


Figura 38 MCUR mantenimiento.  
Fuente: Elaboración propia

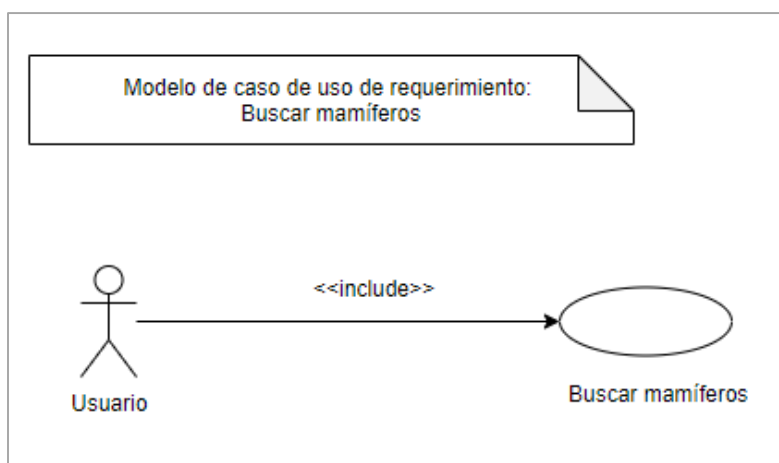


Figura 39 MUCR búsqueda de mamíferos  
Fuente: Elaboración propia

### 8.5.2.6. Especificaciones de casos de uso de requerimiento

#### 8.5.2.6.1. ECUR Seguridad

Tabla 24  
ECUR Seguridad

Partes ECUR	Detalle
Caso de uso	Registra usuario
Actores	Administrador del Sistema
Propósito	Mantener una lista de usuarios registrados
Resumen	Permite mantener una lista de usuarios registrados, que discrimine el acceso al mantenimiento del sistema sólo a aquellos que se encuentran registrados
Pre condición	El Administrador ha presionado el botón “Nuevo” de su interfaz
Flujo básico	<ol style="list-style-type: none"> <li>1. Registrar nuevo usuario</li> <li>2. Buscar usuario</li> <li>3. Modificar un registro, previamente debe haberse seleccionado el que se desee modificar.</li> <li>4. Eliminar un usuario</li> </ol>
Flujo alternativo	<p>Inclusión:</p> <p>A1: Buscar un usuario por su nombre o email de login, y seleccionarlo, de no haber seleccionado uno, mostrar E1.</p>
Excepción	Paso – Acción E1: “Seleccione un registro”.
Post condición	El sistema actualiza la lista de Usuarios

8.5.2.6.2. *ECUR Mantenimiento*

Tabla 25  
*ECUR Mantenimiento: Registrar especie*

Partes ECUR	Detalle
Caso de uso	Registrar mamíferos
Actores	Administrador del sistema
Propósito	Mantener el inventario de mamíferos actualizado
Resumen	Permite mantener el inventario de mamíferos actualizado, que dispone resultados a las búsquedas que los usuarios realizan.
Pre condición	El administrador a presionado el botón “Nuevo” de su interfaz.
Flujo básico	1. Registrar una nueva especie de mamífero
	2. Buscar mamíferos
	3. Modificar un registro, previamente debe haberse seleccionado el que se desee modificar.
	4. Eliminar una especie de mamífero
	5. Referenciar un resultado de especie de mamífero
Flujo alternativo	Inclusión: A1: Buscar un registro de especie de mamífero por su nombre y seleccionarlo, de no haber seleccionado un registro, mostrar E1.
Excepción	Paso – Acción E1: “Seleccione un registro”.
Post condición	El sistema actualiza el registro

Tabla 26

ECUR Mantenimiento: Generar lista invertida

Partes ECUR	Detalle
Caso de uso	Generar lista invertida
Actores	Administrador del sistema
Propósito	Procesar los registros de especies con el algoritmo de índice invertido
Resumen	Permite mantener la lista invertida de registros de tesis, que sirven como base en el resultado de las búsquedas que los usuarios deseen realizar, desde la respectiva interfaz.
Pre condición	El Administrador ha presionado el botón “Generar” de su interfaz
Flujo básico	Generar la Lista Invertida
Flujo alternativo	Inclusión: Ninguna inclusión.
Excepción	Ninguna excepción que el actor tenga que atender.
Post condición	El sistema actualiza la lista invertida.

### 8.5.3. Diagrama de despliegue

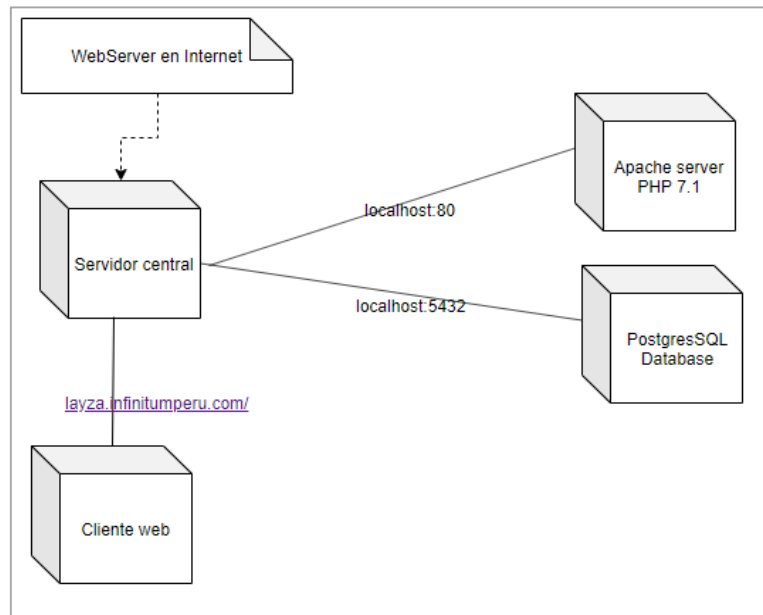


Figura 40 Diagrama de despliegue.  
Fuente: Elaboración propia

#### 8.5.4. Seudocódigo del algoritmo de lista invertida

Input: let D be the set of rows from database

Output: FILE lista\_invertida

N <- size of D

dictionary <- { }

invert\_list <- matrix[]

idf <- matrix[]

FOR EACH document in D

    id <- document[id]

    D <- D - document[id]

    document\_txt <- ""

        FOR EACH attribute in document

            document\_txt <- document\_txt + " " + attribute

        END FOR

    terms <- { }

    DO

        terms <- terms UNION {[first term from document\_txt]}

        document\_txt <- document\_txt - [first term from document\_txt]

    WHILE document\_txt is not empty

    unique\_terms <- matrix[]

        FOR EACH term in terms

            count <- 0

            FOR EACH term1 in terms

                IF term equals term 1 THEN count <- count + 1

            END FOR

            unique\_terms[term] = count

        END FOR

```

FOR EACH term in terms
    dictionary <- dictionary UNION {term}
    invert_list[term][id] <- unique_terms[term]
END FOR
END FOR

FOR EACH term in dictionary
    count_terms <- SIZE OF invert_list[term]
    idf[term] <- log ( N / count_terms )
END FOR

dictionary <- SORT dictionary

content <- {}

FOR EACH term IN dictionary
    data <- matrix[]
    data[term] <- term

    data[id_content] <- {}
    FOR EACH id IN [ KEYS OF invert_list[term] ]
        data[id_content][id] <- invert_list[term][id]
    END FOR

    data[idf] <- idf[term]
    content <- content UNION {data}
END FOR

RETURN FILE lista_invertida WITH CONTENT

```

### 8.5.5. Seudocódigo del algoritmo de búsqueda

```
INPUT:      Let invert_list_content be the content of the inverted list file
            Let terms_query be the set of terms from the query typed in the searchbox
OUTPUT: Set of scored/ranked results

invert_list <- {}
idf <- matrix[]

FOR EACH row IN invert_list_content
    term <- row[term]
    idf[term] <- row[idf]
    i <- 0
    WHILE i LESS OR EQUAL THAN SIZE OF row[id_content]
        invert_list[term][row[id_content][i]] = row[id_content][i+1]
        i <- i + 2
    END WHILE
END FOR

terms <- {}

FOR EACH term IN terms_query
    IF term EXISTS IN KEYS OF invert_list THEN
        terms <- terms UNION {term}
    END IF
END FOR

relevant_document_ids <- {}

FOR EACH term in terms
    ids_documentos <- KEYS OF invert_list[term]
    relevant_document_ids <- relevant_document_ids UNION ids_documentos
END FOR

relevant_document_ids <- relevant_document_ids WITHOUT DUPLICATES

IF SIZE OF relevant_document_ids > 0 THEN
    frecuencia_term_query <- matrix[]

    FOR EACH term in KEYS OF invert_list
        count <- 0
        FOR EACH term1 in terms
            IF term equals term 1 THEN count <- count + 1
        END FOR

        frecuencia_term_query[term] = count
    END FOR
```

```

query_norm <- SQUARE ROOT OF k

weight <- matrix[]

FOR EACH term IN SET OF KEYS FROM invert_list
  FOR EACH id IN relevant_document_ids
    IF id EXISTS IN SET OF KEYS FROM invert_list[term] THEN
      weight[id][term] <- invert_list[term][id] * idf[term]
    ELSE
      weight[id][term] <- 0
    END IF
  END FOR
END FOR
scores <- matrix[]

FOR EACH id IN relevant_document_ids
  document_norm <- 0
  FOREACH term IN SET OF KEYS FROM invert_list
    document_norm <- document_norm + weight[id][term] ^ 2
  END FOR

  div <- ( SQUARE ROOT OF document_norm ) * query_norm

  IF div NOT EQUALS 0 THEN
    sum_value_coordenada <- 0
    FOR EACH term IN SET OF KEYS FROM invert_list
      sum_value_coordenada <- sum_value_coordenada +
weight[id][term] * weight_query[term]
    END FOR

    document_norm <- 0
    FOREACH term IN SET OF KEYS FROM invert_list
      document_norm <- document_norm + weight[id][term] ^ 2
    END FOR

    document_norm <- SQUARE ROOT OF document_norm

    scores[id] <- sum_value_coordenada / ( document_norm * query_norm )
  ELSE
    scores[id] <- 1
  END IF
END FOR

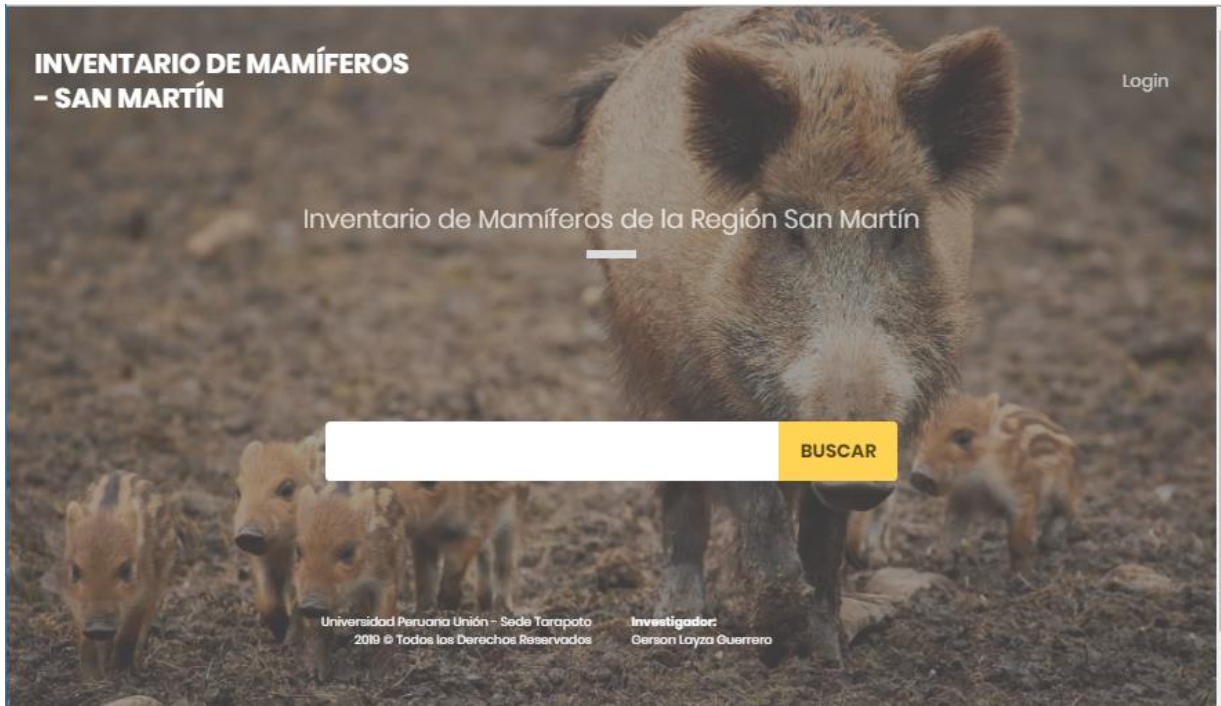
scores <- scores INVERSE SORTED
END IF

RETURN scores

```



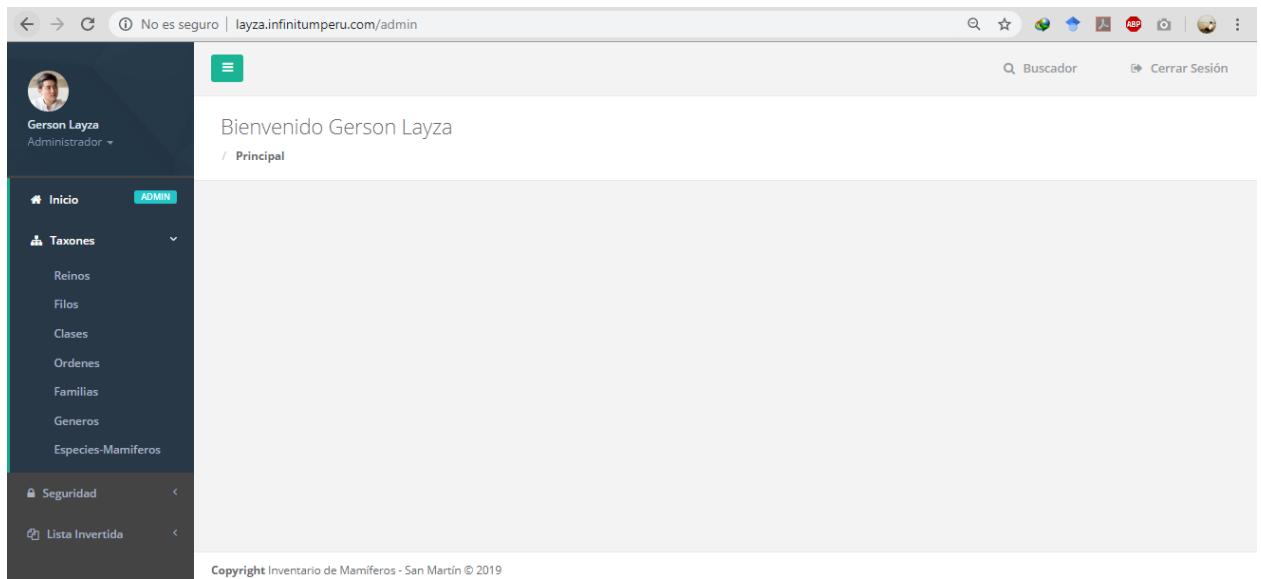
## 8.6. Pantallazos del sistema



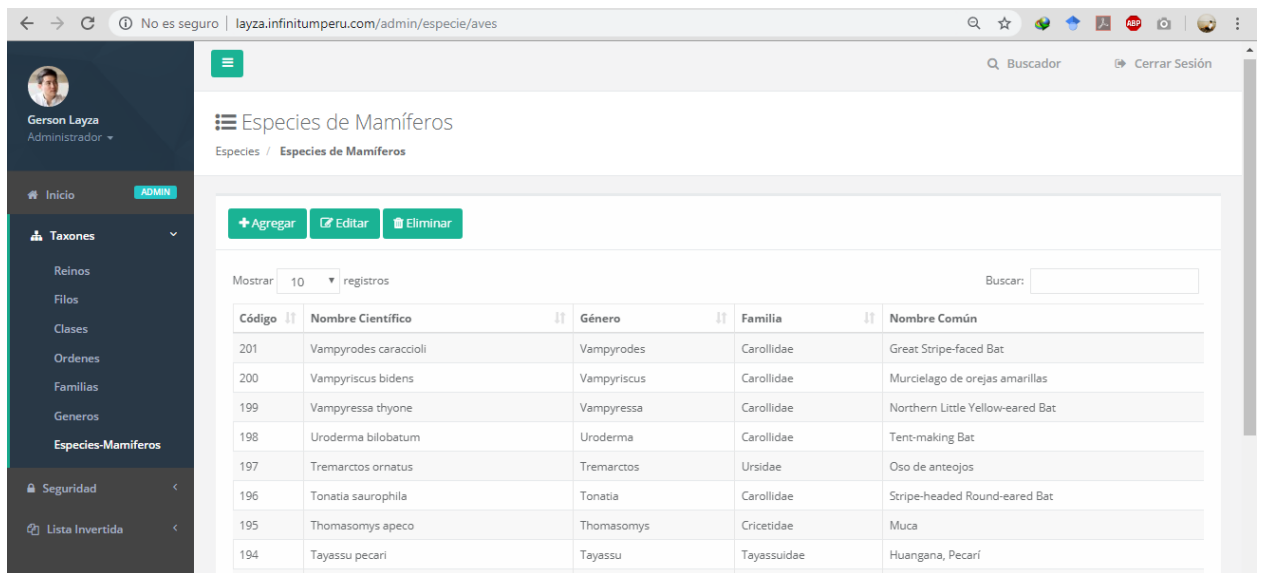
Anexos 10 Página principal del motor de búsqueda de mamíferos



Anexos 11 Resultados de búsqueda



*Anexos 12 Panel de control del motor de búsqueda*



*Anexos 13 Inventario de especie de mamíferos*

## **8.7. Manual del usuario**

**Universidad Peruana Unión**  
**Facultad de Ingeniería y Arquitectura**  
**Escuela Profesional de Ingeniería de**  
**Sistemas**



*Una Institución Adventista*

**Manual de Usuario del**  
**Motor de Búsqueda Mamíferos San Martín**

Morales, junio 2019

### **8.7.1. Introducción**

El propósito de este manual es facilitar al usuario el uso del motor de búsqueda Mamíferos San Martín a través de diferentes instrucciones, pasos, y captura de pantallas que se presentarán en esta sección. Así mismo, se pretende mostrar la funcionalidad de los módulos con los que cuenta el mismo.

El uso del motor de búsqueda, ver *Figura 41*, está disponible para todo el público y comunidad científica. Sin embargo, es importante mencionar, que el módulo de administrador del sistema es de uso de exclusivo del investigador, y los taxónomos expertos en el área que gestionan la información científica de mamíferos.

## **8.7.2. Requerimientos técnicos**

### **8.7.2.1. *Requerimientos mínimos de Hardware***

- Procesador: Core
- Memoria RAM: Mínimo: 1 Gigabytes (GB)
- Disco Duro: 500Gb.

### **8.7.2.2. *Requerimientos mínimos de Software***

- Sistema Operativo: Windows v7

## **8.7.3. Herramientas utilizadas para el desarrollo**

### **8.7.3.1. *PHP***

Es un clásico lenguaje de programación muy usado por público. Inicialmente fue creado para el desarrollo web de contenido ágil y dinámico. Está actualmente entre los proyectos de código abierto más populares (gracias en parte a la similitud de su sintaxis con el lenguaje C).

Las ejecuciones son realizadas en el Servidor y el cliente es el encargado de recibir los resultados de la ejecución. Si el cliente realiza una petición, se ejecuta el intérprete de PHP y se genera el contenido de manera dinámica. Permite conexión con varios tipos de Bases de Datos como: MySQL, Oracle, Postgress, SQL Server, etc. permitiendo aplicaciones robustas sobre la WEB. Este lenguaje de programación puede ser ejecutado en la gran mayoría de sistemas operacionales y puede interactuar con Servidores WEB populares.

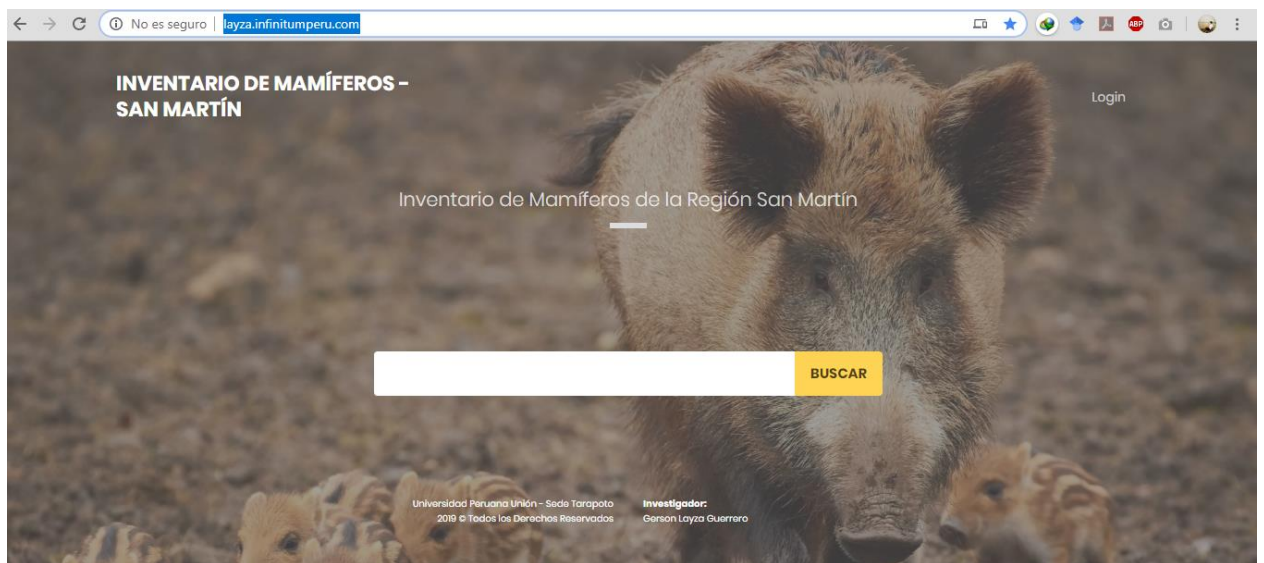
### **8.7.3.2. *PostgreSQL***

Es un sistema que permite gestionar bases de datos de tipo relacional orientado a objetos y de código abierto, el nombre de la licencia es PostgreSQL. Como muchos otros proyectos de código abierto, el desarrollo de PostgreSQL está dirigido por una comunidad de desarrolladores que trabajan de forma desinteresada, altruista, libre o

patrocinados por instituciones comerciales. La comunidad es nombrada como PGDG (PostgreSQL Global Development Group).

#### 8.7.4. Funcionalidad principal

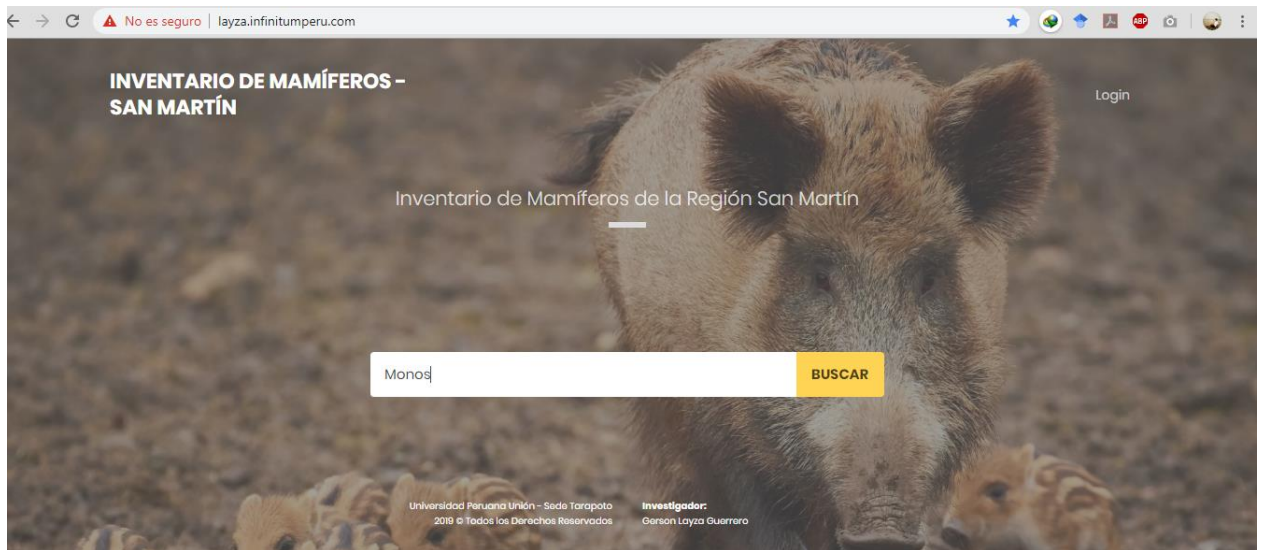
En un navegador popular ingrese al siguiente link, <http://layza.infinitumperu.com/> inmediatamente tendrán acceso a la interfaz principal, que es el buscador.



*Figura 41 Página principal del buscador*

#### 8.7.4.1. *Ingresando al sistema*

En la funcionalidad hay que tener en cuenta que el sistema es un buscador, y la información con la que cuenta son inventarios de mamíferos de la región San Martín, que fueron organizados y procesados para su posterior presentación. Para ver la funcionalidad de sistema solo hay que ingresar correctamente el nombre del mamífero que deseamos saber en la barra blanca del buscador. Ver figura...



*Figura 42 Ingreso valores para búsqueda*

Luego de ingresar el nombre, en nuestro caso ingresamos “monos”, le damos clic en el botón amarillo buscar.



*Figura 43 Ejecución de búsqueda*



Esperamos a ver, a que el buscador haga su trabajo y en cuestión de milésimas de segundos nos retorne una lista de resultados que se asemejan a nuestra consulta.

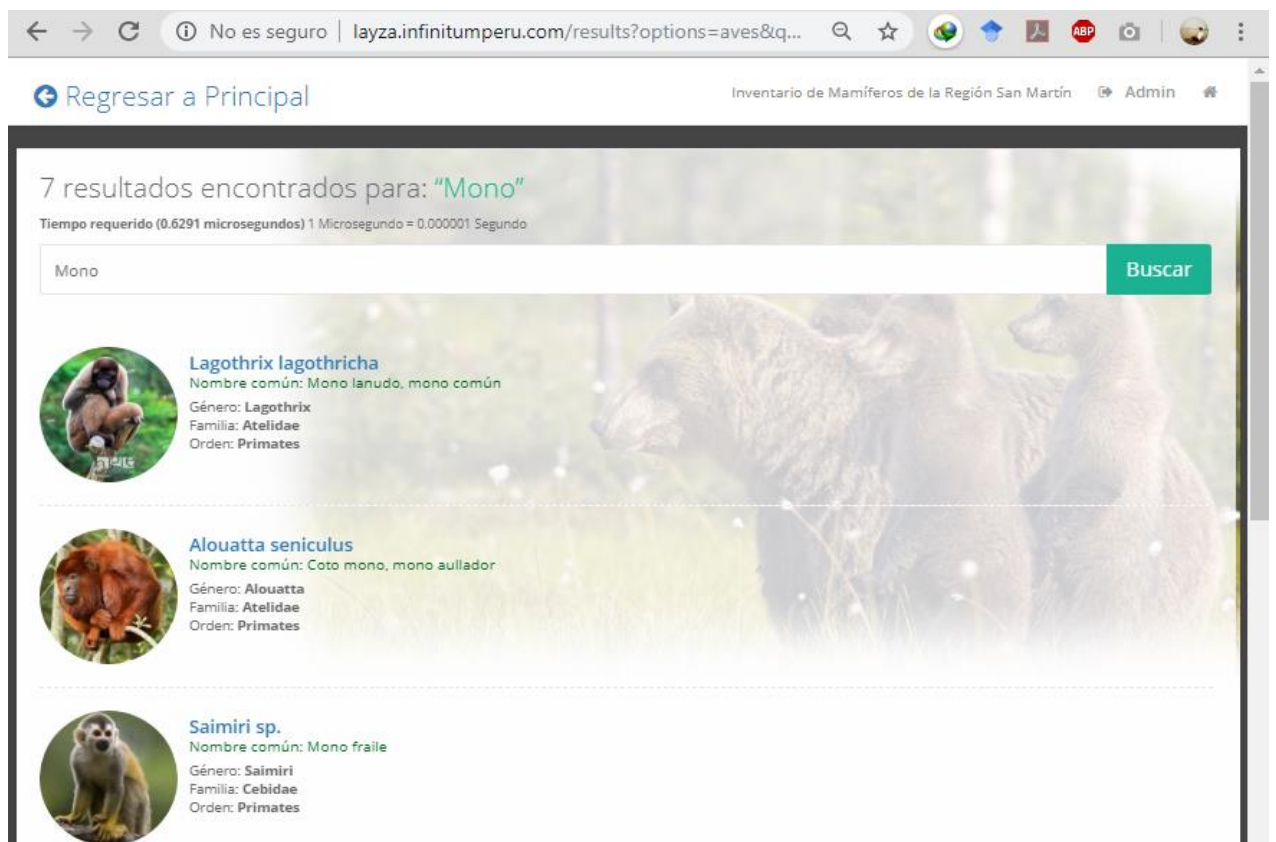


Figura 44 Resultados de búsqueda

## 8.7.5. Funcionalidad de administrador

### 8.7.5.1. Ingreso al módulo administrador

El acceso al panel de administrador es únicamente proporcionado por el profesional técnico. Para el ingreso al módulo, hay que hacer clic en "Login", que se encuentra en parte superior derecha.

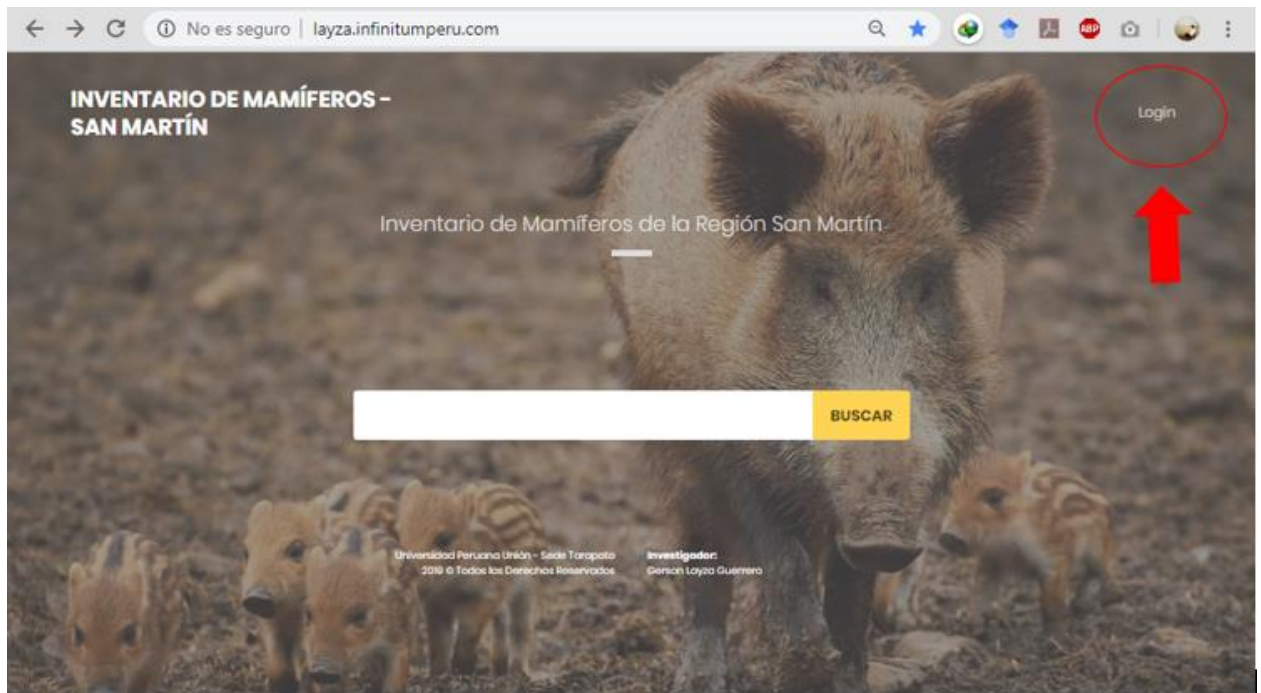


Figura 45 Ingreso al módulo administrador

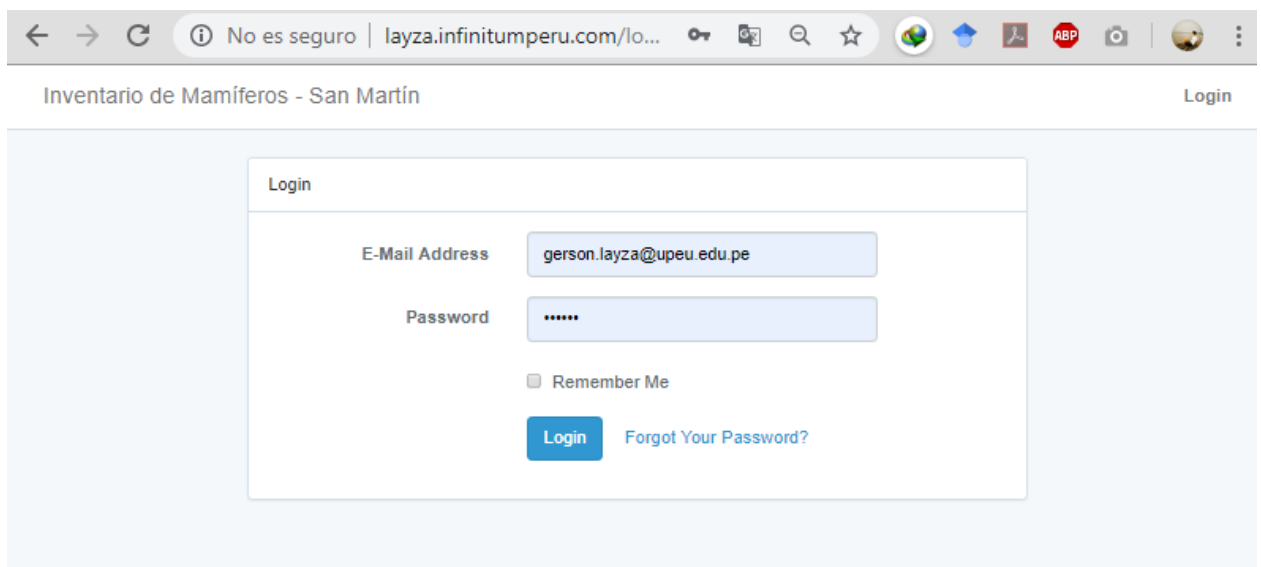


Figura 46 Inicio de sesión al módulo administrador

### 8.7.6. Interfaz del administrador

La siguiente sección muestra los módulos de administración del sistema. La ejecución interna de estos procesos permite que el buscador tenga funcionalidad. En el menú principal las pantallas del sistema se dividen en 3 zonas:

- a. Zona de título
- b. Zona de Índice o menú

c. Zona de área de trabajo

The screenshot shows a web application interface for an administrator. The browser address bar displays 'layza.infinitemperu.com/admin/reinos#'. The user is identified as 'Gerson Layza, Administrador'. The main content area is titled 'Reino' and 'Taxones / Reino'. It features a sidebar menu with options like 'Inicio', 'Taxones', 'Reinos', 'Filos', 'Clases', 'Ordenes', 'Familias', 'Generos', 'Especies-Mamiferos', 'Seguridad', and 'Lista Invertida'. The main area has buttons for '+ Agregar', 'Editar', and 'Eliminar'. Below these are controls for 'Mostrar 10 registros' and a search box. A table lists two records:

Código	Descripción
7	Plantae
5	Animalia

At the bottom, it indicates 'Mostrando registros del 1 al 2 de un total de 2 registros' and includes navigation buttons 'Anterior', '1', and 'Siguiente'. Red annotations are present: 'A' points to the 'Reino' header, 'B' circles the sidebar menu, and 'C' points to the 'Eliminar' button.

Figura 47 Modulo del administrador

### 8.7.6.1. Barras y botones

La forma de utilizar esta herramienta es la misma que en cualquier aplicación que use Windows, también puede usar los comandos del teclado para seleccionar opciones, así como la funcionalidad a través del ratón.

Es importante mencionar que en casi todos módulos del sistema, los botones y las barras pueden estar visibles o no, dependiendo de las acciones a ejecutar.

Los botones principales que se presentan a continuación, se encuentran dentro de la zona de área de trabajo, en la parte superior, ver

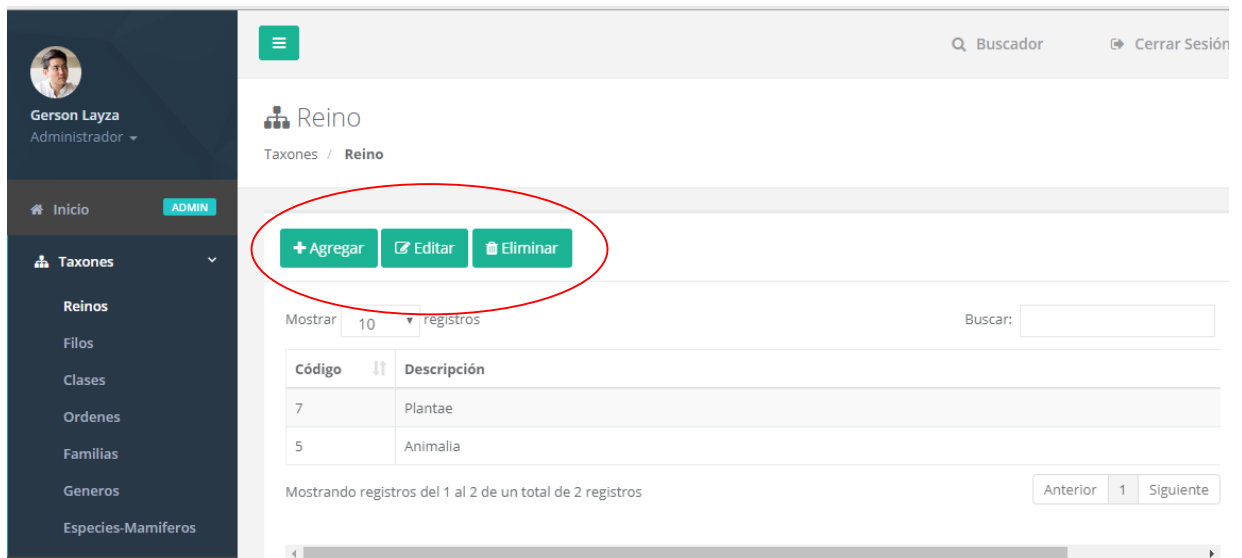


Figura 48 Botones principales

Tabla 27 Botones principales

#	Botones	Descripción
1		Este botón permite iniciar un nuevo registro de datos. Ver <i>Figura 52</i>
2		Este botón permite modificar el registro seleccionado. Ver <i>Figura 50</i>
3		Permite eliminar el registro seleccionado. Ver <i>Figura 51</i>
4		Guarda o Almacena el registro. Ver <i>Figura 49</i>

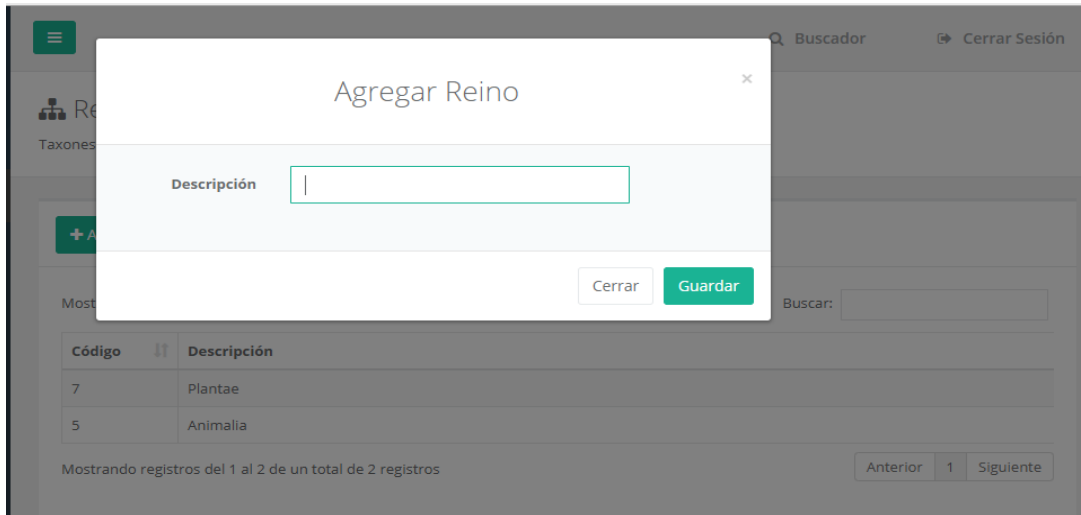


Figura 49 Botón Guardar

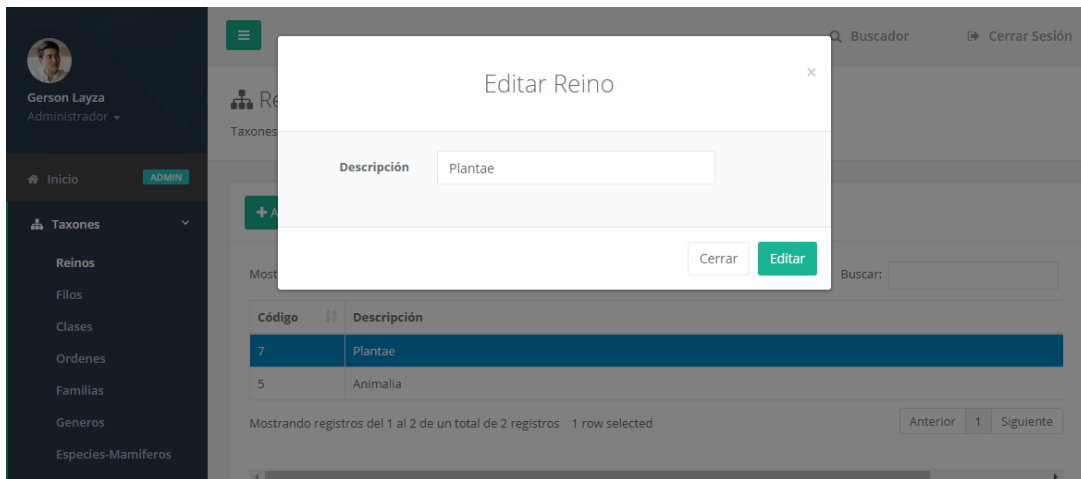


Figura 50 Botón editar

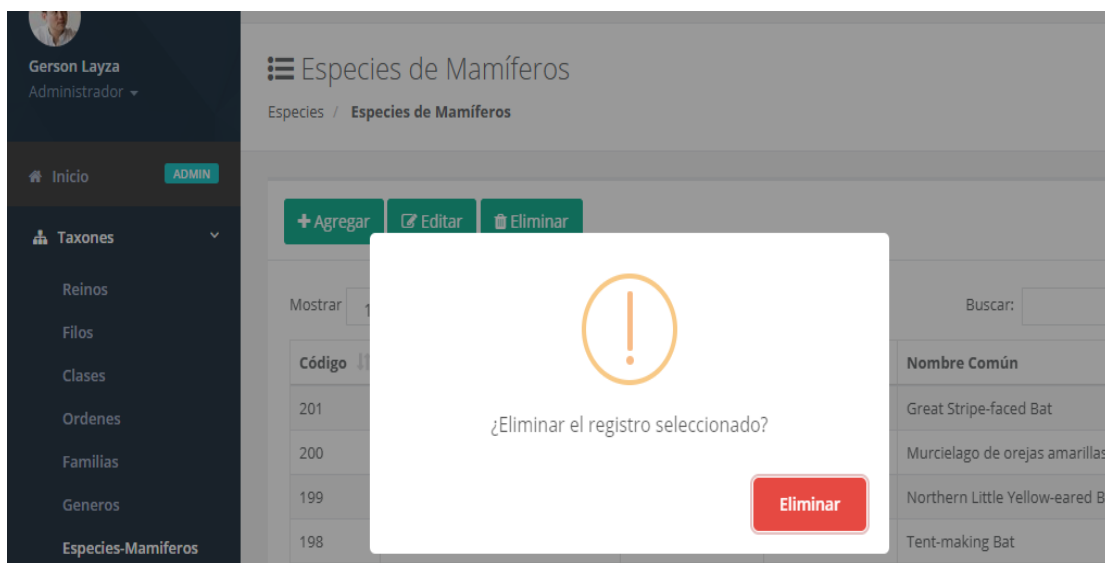


Figura 51 Botón eliminar

← → ↻ No es seguro | layza.infinitemperu.com/admin/especie/aves

Gerson Layza Administrador

Inicio ADMIN

Taxones

- Reinos
- Filos
- Clases
- Órdenes
- Familias
- Generos
- Especies-Mamíferos**

Seguridad

Lista Invertida

### Especies de Mamíferos

Especies / Especies de Mamíferos

**+ Agregar** **Editar** **Eliminar**

Mostrar 10 registros

Buscar:

Código	Nombre Científico	Género	Familia	Nombre Común
201	Vampyroides caraccioli	Vampyroides	Carollidae	Great Stripe-faced Bat
200	Vampyriscus bidens	Vampyriscus	Carollidae	Murciélago de orejas amarillas
199	Vampyressa thyone	Vampyressa	Carollidae	Northern Little Yellow-eared Bat
198	Uroderma bilobatum	Uroderma	Carollidae	Tent-making Bat
197	Tremarctos ornatus	Tremarctos	Ursidae	Oso de anteojos
196	Tonatia saurophila	Tonatia	Carollidae	Stripe-headed Round-eared Bat
195	Thomasomys apeco	Thomasomys	Cricetidae	Muca
194	Tayassu pecari	Tayassu	Tayassuidae	Huangana, Pecarí
193	Tapirus terrestris	Tapirus	Tapiridae	Sachavaca
192	Tapirus pinchaque	Tapirus	Tapiridae	Sachavaca, danta cordillerana

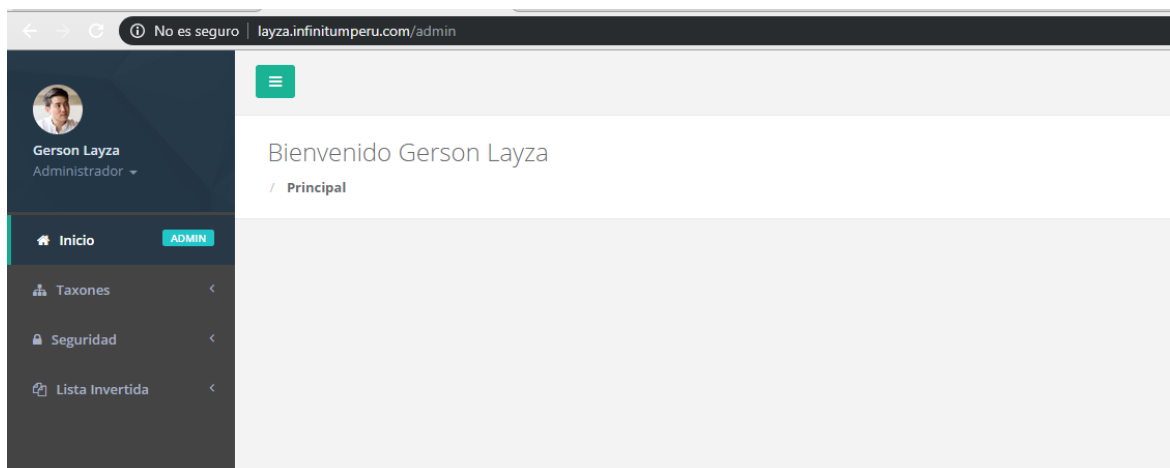
Mostrando registros del 1 al 10 de un total de 123 registros

Anterior 1 2 3 4 5 ... 13 Siguiente

Figura 52 Botón agregar

### 8.7.7. Módulos que Integran el sistema

Los Módulos que integran el sistema se encuentran en la zona del índice o menú del lado izquierdo, como se muestra a continuación:

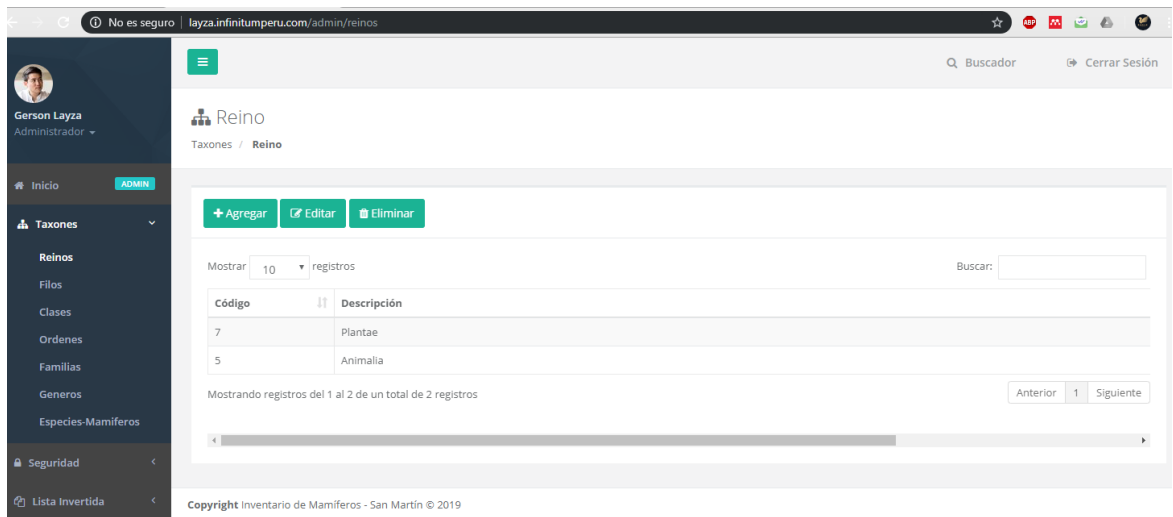


*Figura 53 Vista principal - Modulo administrador*

#### 8.7.7.1. Taxones

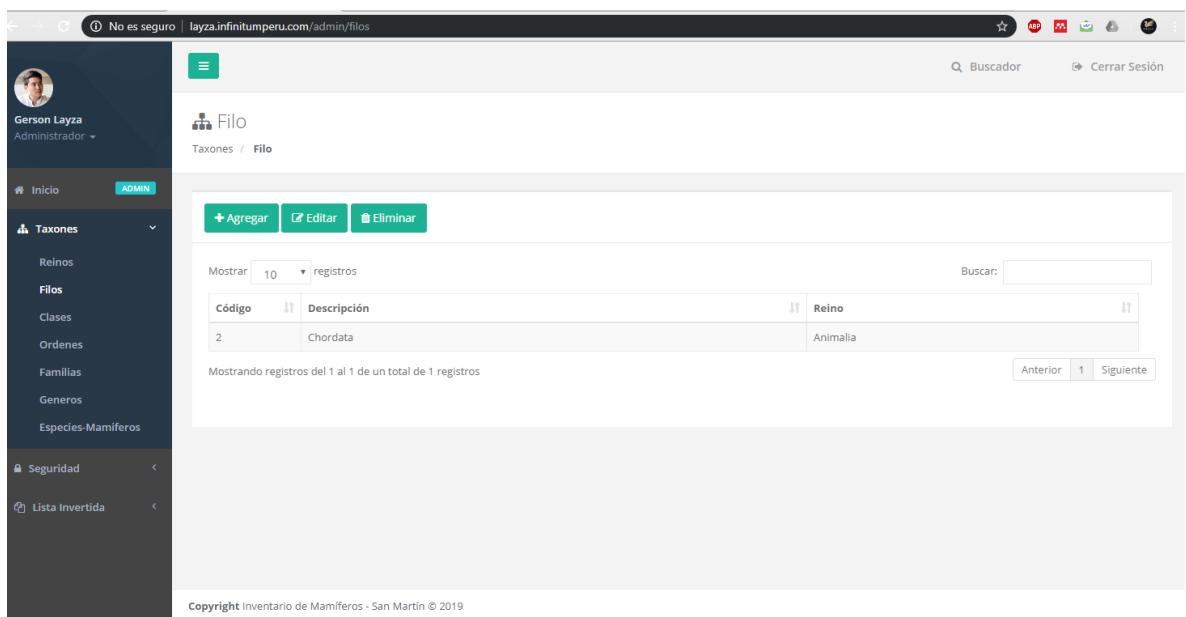
En módulo permite organizar un conjunto de organismos emparentado, que en una clasificación científica realizada han sido agrupados. En este procedimiento los taxones agregan un nombre en latín, una descripción si es una especie, una clase, un género.

- **Reinos:** también conocido en latín como Animalia (animal) o Metazoos (metazoa), es un conjunto de seres vivos que comparten características relevantes que los distingue de otros.



*Figura 54 Taxones - reino*

- **Filos:** El filo es la subdivisión básica del Reino animal y puede definirse como una agrupación de animales basada en su plan general de organización. Contiene a la agrupación de animales según su descripción y reino.



*Figura 55 Taxones - filos*



- Clases: Es la clasificación que agrupa a los animales según su descripción y reino.

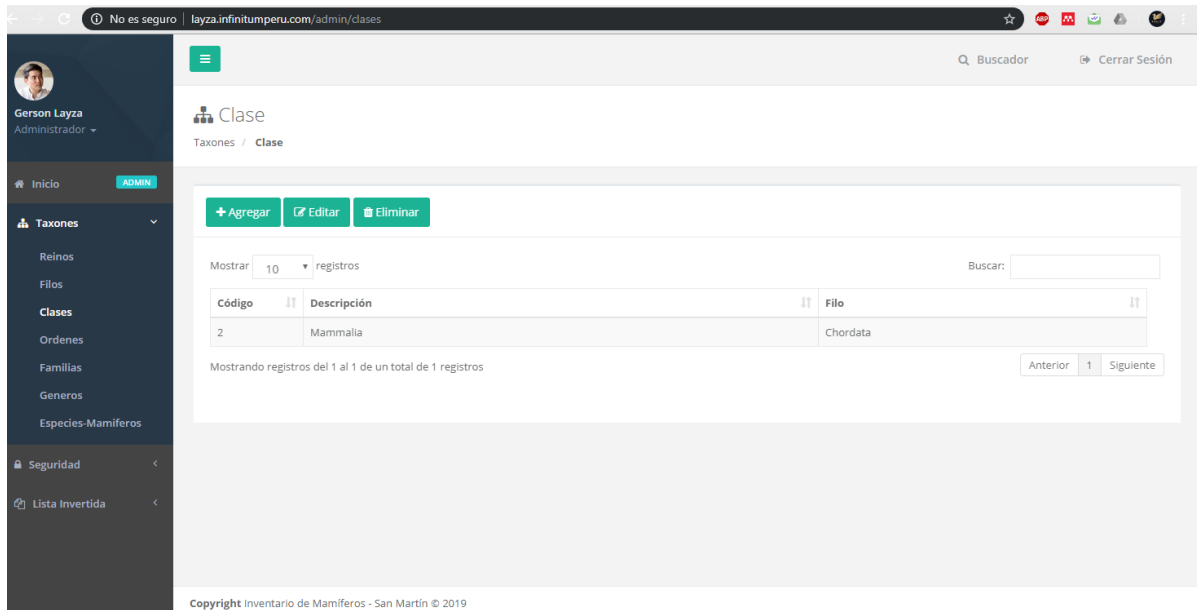


Figura 56 Taxones - clases

- Ordenes: Es la clasificación que agrupa a los animales según su descripción, clase y filo.

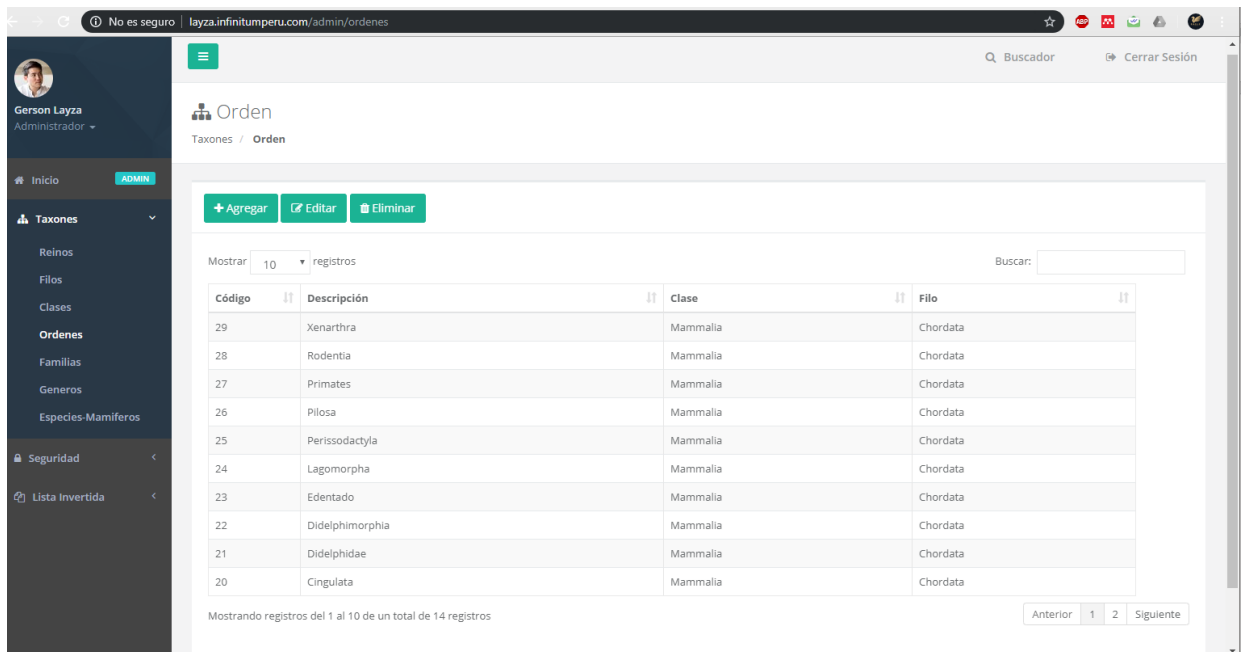


Figura 57 Taxones - ordenes

- Familias: Es la clasificación que agrupa a los animales según su descripción, orden, clase, y filo.

Mostrar 10 registros

Código	Descripción	Orden	Clase	Filo
44	Agoutidae	Rodentia	Mammalia	Chordata
45	Aotidae	Primates	Mammalia	Chordata
46	Atelidae	Primates	Mammalia	Chordata
47	Bradypodidae	Xenarthra	Mammalia	Chordata
48	Bradypodidae	Pilosa	Mammalia	Chordata
49	Callitrichidae	Primates	Mammalia	Chordata
50	Canidae	Carnivora	Mammalia	Chordata
51	Carollidae	Chiroptera	Mammalia	Chordata
52	Caviidae	Rodentia	Mammalia	Chordata
53	Cebidae	Primates	Mammalia	Chordata

Mostrando registros del 1 al 10 de un total de 42 registros

Figura 58 Taxones - familias

- Géneros: Clasificación que agrupa a los animales según su descripción, familia, y orden.

Mostrar 10 registros

Código	Descripción	Familia	Orden
315	Vampyrodes	Carollidae	Chiroptera
314	Vampyriscus	Carollidae	Chiroptera
313	Vampyressa	Carollidae	Chiroptera
312	Uroderma	Carollidae	Chiroptera
311	Tremarctos	Ursidae	Carnivora
310	Tonatia	Carollidae	Chiroptera
309	Thomasomys	Cricetidae	Rodentia
308	Tayassu	Tayassuidae	Artiodactyla
307	Tapirus	Tapiridae	Perissodactyla
306	Tamandua	Myrmecophagidae	Edentado

Mostrando registros del 1 al 10 de un total de 92 registros

Figura 59 Taxones – géneros

- Especies: Clasificación que agrupa a los animales según su nombre científico, género, familia, y nombre común.

Código	Nombre Científico	Género	Familia	Nombre Común
201	Vampyrodes caraccioli	Vampyrodes	Carollidae	Great Stripe-faced Bat
200	Vampyriscus bidens	Vampyriscus	Carollidae	Murcielago de orejas amarillas
199	Vampyressa thylene	Vampyressa	Carollidae	Northern Little Yellow-eared Bat
198	Uroderma bilobatum	Uroderma	Carollidae	Tent-making Bat
197	Tremarctos ornatus	Tremarctos	Ursidae	Oso de anteojos
196	Tonatia saurophila	Tonatia	Carollidae	Stripe-headed Round-eared Bat
195	Thomasomys apeco	Thomasomys	Cricetidae	Muca
194	Tayassu pecari	Tayassu	Tayassuidae	Huangana, Pecarí
193	Tapirus terrestris	Tapirus	Tapiridae	Sachavaca
192	Tapirus pinchaque	Tapirus	Tapiridae	Sachavaca, danta cordillerana

Figura 60 Taxones - especies de mamíferos

### 8.7.7.2. Seguridad

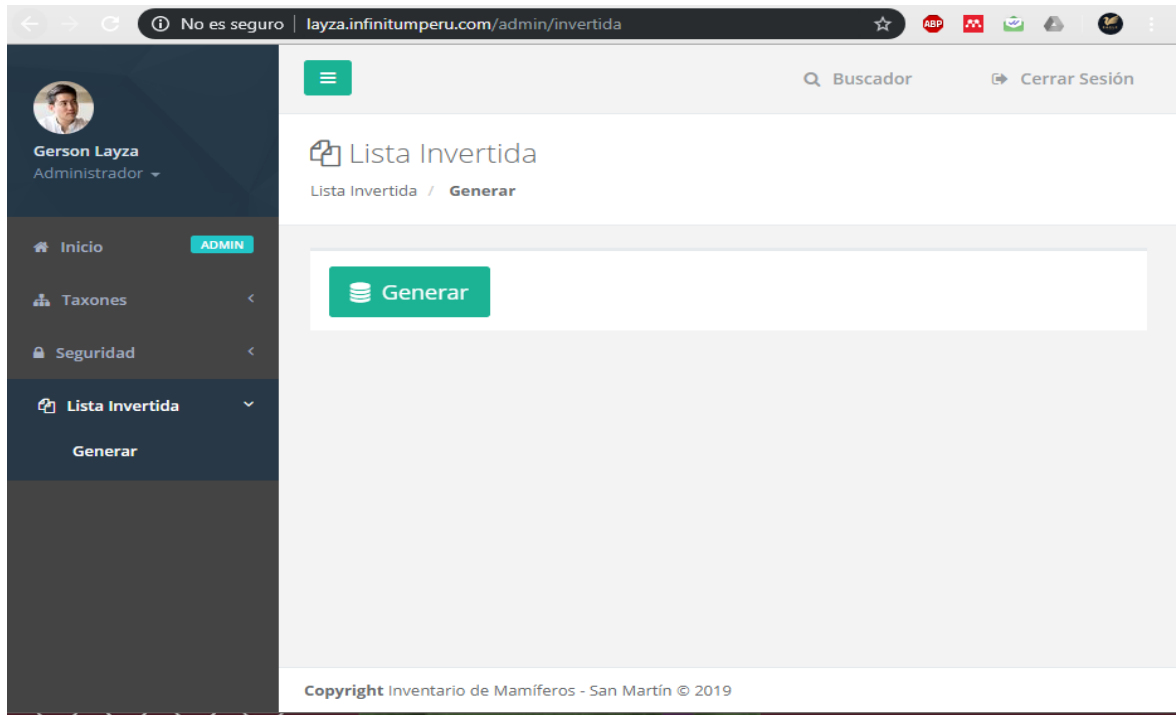
Este módulo permite registrar y dar accesos a los usuarios, para la administración del sistema. Los usuarios idóneos para administración del sistema son los taxónomos.

Rendering engine	Browser	Platform(s)	Engine version	CSS grade
23	qwer	asdfadsf@adsfasd.com	2017-10-28 22:50:28	2017-10-28 22:50:28
22	prueba	qwerqwer@asfasdf.com	2017-10-28 22:34:40	2017-10-28 22:34:40
21	advrucoba@yahoo.es	advrucoba@yahoo.es	2017-10-28 22:13:51	2017-10-28 22:13:51
3	Pedro Gonzales	pedro.gonzales0306@gmail.com	2019-08-22 21:03:20	2019-08-22 21:03:20
2	Gerson Layza	gerson.layza@upeu.edu.pe	2019-01-28 21:47:04	2019-01-28 21:47:04

Figura 61 Módulo de seguridad

### 8.7.7.3. *Lista invertida*

Este módulo permite realizar el procedimiento principal, para el funcionamiento del motor de búsqueda. El acceso está disponible para los técnicos quienes son los encargados de generar la lista invertida, este procedimiento será realizado en coordinación con los taxónomos.



*Figura 62 Módulo de lista invertida*