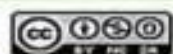




Esta obra está bajo una [Licencia
Creative Commons Atribución-
NoComercial-Compartirigual 2.5 Perú](http://creativecommons.org/licenses/by-nc-sa/2.5/pe/).

Vea una copia de esta licencia en
<http://creativecommons.org/licenses/by-nc-sa/2.5/pe/>



UNIVERSIDAD NACIONAL DE SAN MARTÍN- TARAPOTO
FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA



Buscador informático basado en el modelo vectorial para la difusión de información científica de las Especies de palmeras (Palmae) inventariadas de la Región San Martín

Tesis para optar al título profesional de Ingeniero de Sistemas e Informática

AUTOR:

Carolina Mendoza Amasifuen

ASESOR:

Ing. M.Sc. Pedro Antonio Gonzales Sánchez

Tarapoto - Perú

2018

UNIVERSIDAD NACIONAL DE SAN MARTÍN -TARAPOTO
FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E
INFORMÁTICA



Buscador informático basado en el modelo vectorial para la difusión de información científica de las Especies de palmeras (Palmae) inventariadas de la Región San Martín

AUTOR:

Carolina Mendoza Amasifuen.


Sustentada y aprobada el día 05 de diciembre del 2018, por los siguientes jurados:



Ing. M. Sc. Pamela Magnolia Granda Milón
Presidente



Ing. M. Sc. José Enrique Celis Escudero
Secretario



Ing. M. Sc. Andy Hirvyn Rucoba Reátegui
Miembro

Declaratoria de Autenticidad

Yo, Carolina Mendoza Amasifuen, bachiller en Ingeniería de Sistema e Informática, de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de San Martín - Tarapoto, identificado con DNI N°48428946, con la tesis titulada “Buscador informático basado en el modelo vectorial para la difusión de información científica de las Especies de palmeras (Palmae) inventariadas de la Región San Martín”

Declaro bajo juramento que:

1. La tesis presentada es de mi autoría.
2. He respetado las normas internacionales de citas y referencias para las fuentes consultadas. Por lo tanto, la tesis no ha sido plagiada ni total ni parcialmente.
3. La tesis no ha sido plagiada, es decir, no ha sido publicada ni presentada anteriormente para obtener algún grado académico previo o título profesional.
4. Los datos presentados en los resultados reales, no han sido falsificados, ni duplicados, ni copiados y por lo tanto los resultados que se presenten en la tesis se constituirán en aportes a la realidad investigada.

De considerar cuenta con una falta grave, como el hecho de contar con datos fraudulentos, demostrar indicios y plagio (al no citar información de sus autores). Plagio (al presentar información de otros trabajos como propios), falsificación de (al presentar la información de ideas de otras personas de forma falsa), entre otros, asumo las consecuencias y sanciones que de mi acción se deriven, sometiéndome a la normatividad vigente de la Universidad Nacional de San Martín – Tarapoto.

Tarapoto, 05 de diciembre del 2018



Carolina Mendoza Amasifuen

DNI N° 48428946



Formato de autorización NO EXCLUSIVA para la publicación de trabajos de investigación, conducentes a optar grados académicos y títulos profesionales en el Repositorio Digital de Tesis.

1. Datos del autor:

Apellidos y nombres:	Mendoza Amasifuen Carolina		
Código de alumno :	127141	Teléfono:	928187155
Correo electrónico :	Carolina061294@gmail.com	DNI:	98428946

(En caso haya más autores, llenar un formulario por autor)

2. Datos Académicos

Facultad de:	Ingeniería de Sistemas e Informática
Escuela Profesional de:	Ingeniería de Sistemas e Informática

3. Tipo de trabajo de investigación

Tesis	(X)	Trabajo de investigación	()
Trabajo de suficiencia profesional	()		

4. Datos del Trabajo de investigación

Título:	Buscador informático basado en el modelo vectorial para la difusión de información científica de las especies de palmeras (palmae) inventariadas de la región San Martín
Año de publicación:	2018

5. Tipo de Acceso al documento

Acceso público *	(X)	Embargo	()
Acceso restringido **	()		

Si el autor elige el tipo de acceso abierto o público, otorga a la Universidad Nacional de San Martín – Tarapoto, una licencia **No Exclusiva**, para publicar, conservar y sin modificar su contenido, pueda convertirla a cualquier formato de fichero, medio o soporte, siempre con fines de seguridad, preservación y difusión en el Repositorio de Tesis Digital. Respetando siempre los Derechos de Autor y Propiedad Intelectual de acuerdo y en el Marco de la Ley 822.

En caso que el autor elija la segunda opción, es necesario y obligatorio que indique el sustento correspondiente:

6. Originalidad del archivo digital.

Por el presente dejo constancia que el archivo digital que entrego a la Universidad Nacional de San Martín - Tarapoto, como parte del proceso conducente a obtener el título profesional o grado académico, es la versión final del trabajo de investigación sustentado y aprobado por el Jurado.

7. Otorgamiento de una licencia *CREATIVE COMMONS*

Para investigaciones que son de acceso abierto se les otorgó una licencia *Creative Commons*, con la finalidad de que cualquier usuario pueda acceder a la obra, bajo los términos que dicha licencia implica

<https://creativecommons.org/licenses/by-nc-sa/2.5/pe/>

El autor, por medio de este documento, autoriza a la Universidad Nacional de San Martín - Tarapoto, publicar su trabajo de investigación en formato digital en el Repositorio Digital de Tesis, al cual se podrá acceder, preservar y difundir de forma libre y gratuita, de manera íntegra a todo el documento.

Según el inciso 12.2, del artículo 12° del Reglamento del Registro Nacional de Trabajos de Investigación para optar grados académicos y títulos profesionales - RENATI **“Las universidades, instituciones y escuelas de educación superior tienen como obligación registrar todos los trabajos de investigación y proyectos, incluyendo los metadatos en sus repositorios institucionales precisando si son de acceso abierto o restringido, los cuales serán posteriormente recolectados por el Repositorio Digital RENATI, a través del Repositorio ALICIA”**.

Firma del Autor

8. Para ser llenado en la Oficina de Repositorio Digital de Ciencia y Tecnología de Acceso Abierto de la UNSM – T.

Fecha de recepción del documento:

08, 05, 2019



Firma del Responsable de Repositorio Digital de Ciencia y Tecnología de Acceso Abierto de la UNSM – T.

***Acceso abierto:** uso lícito que confiere un titular de derechos de propiedad intelectual a cualquier persona, para que pueda acceder de manera inmediata y gratuita a una obra, datos procesados o estadísticas de monitoreo, sin necesidad de registro, suscripción, ni pago, estando autorizada a leerla, descargarla, reproducirla, distribuirla, imprimirla, buscarla y enlazar textos completos (Reglamento de la Ley No 30035).

** **Acceso restringido:** el documento no se visualizará en el Repositorio.

Dedicatoria

Dedico esta Tesis a toda mi familia, especialmente a mis padres Manuel Mendoza Pinchi y Alcidia Amasifuen Gonzales, por haberme apoyado en todo momento, por sus consejos, sus valores, por la motivación constante y fueron ellos mi impulso para culminar con éxito esta etapa de mi vida.

A mis hermanas por estar siempre presentes, acompañándome y por el apoyo moral, que me brindaron.

Agradecimiento

A Dios por haberme permitido llegar hasta este punto y haberme dado salud para lograr mis objetivos.

A mis padres por su gran esfuerzo por darme un futuro mejor, por impulsar el desarrollo de mi formación profesional y porque siempre estuvieron conmigo brindándome su apoyo económico y emocional incondicionalmente.

Al Ing M.Sc. Pedro Antonio Gonzales Sánchez por su gran apoyo y motivación para la elaboración de esta tesis, por su tiempo compartido y por haber confiado en mí y haberme animado a emprender la elaboración de esta tesis.

Al Blgo. Ángel Martín Rodríguez del Castillo, por haberme abierto las puertas y compartir sus conocimientos conmigo.

Por último, gracias a todas las personas que me han acompañado y animado en este largo camino.

A todos ustedes mi mayor reconocimiento y gratitud.

Índice

Dedicatoria.....	vi
Agradecimiento.....	vii
Índice.....	viii
Índice de figuras.....	x
Índice de tablas.....	xi
Listas de siglas y abreviaturas.....	xii
Resumen.....	xiii
Abstract.....	xiv
INTRODUCCIÓN.....	1
CAPÍTULO I.....	3
REVISIÓN BIBLIOGRÁFICA.....	3
1.1. Fundamento teórico científico.....	3
1.1.1. Buscador Informático.....	3
1.1.2. Modelos para la recuperación de información.....	4
1.1.3. Modelo de recuperación vectorial.....	5
1.1.4. Difusión de información.....	7
1.1.5. Palmeras.....	11
1.2. Definición de términos básicos.....	14
1.2.1. Especie.....	14
1.2.2. Información científica.....	14
1.2.3. Inventario.....	14
CAPÍTULO II.....	16
MATERIALES Y MÉTODOS.....	16
2.1. Sistema de hipótesis.....	16
2.2. Sistema de variables.....	16
2.3. Tipo de método de la investigación.....	17
2.4. Diseño de investigación.....	18

2.5. Población y muestra	18
2.6. Técnicas de recolección de datos	19
2.7. Técnicas de procedimiento	19
CAPÍTULO III.....	21
RESULTADOS DE LA INVESTIGACIÓN.....	21
3.1. Presentación de resultados.....	21
3.2. Contratación de hipótesis	32
3.2. Discusión de resultados	35
CONCLUSIONES.....	36
RECOMENDACIONES	37
BIBLIOGRAFÍA	38
ANEXOS	40

Índice de figuras

Figura 1: Fórmula del Modelo Vectorial.....	5
Figura 2: Cálculo del TF de un término	6
Figura 3: Fórmula del factor IDF.....	7
Figura 4: Fórmula MAP	7
Figura 5: Difusión Documental	9
Figura 6: Vía de Difusión de la información Pre Test	22
Figura 7: Forma de Difusión de la información Pre Test.....	23
Figura 8: Nivel de Difusión de la información Pre Test	24
Figura 9: Vía de Difusión de la información Post Test	27
Figura 10: Forma de Difusión de la información Post Test	28
Figura 11: Nivel de Difusión de la información Post Test.....	29
Figura 12: Similitud Consulta- Respuesta Post Test.....	30
Figura 13: Satisfacción del usuario Post Test.....	31
Figura 14: Nivel de difusión de información Pre y Post implementación del buscador informático	32
Figura 15: Región Crítica Distribución T-Student.....	34

Índice de tablas

Tabla 1 Escala de medición de las variables	17
Tabla 2 Técnicas e instrumentos a utilizar en el estudio	19
Tabla 3 Resumen de los indicadores de la difusión de información, antes de la implementación del buscador.....	21
Tabla 4 Vía de difusión Pre Test	22
Tabla 5 Forma de difusión Pre Test	23
Tabla 6 Nivel de difusión Pre Test	24
Tabla 7 Resumen de los indicadores de la difusión de información, después de la implementación del buscador informático	25
Tabla 8 Vía de difusión Post Test	26
Tabla 9 Forma de difusión Post Test	27
Tabla 10 Nivel de difusión Post Test.....	28
Tabla 11 Similitud Consulta-Respuesta Post Test.....	29
Tabla 12 Satisfacción del usuario Post Test	30
Tabla 13 Comparación proporcional del nivel de difusión de información en relación con la implementación del buscador informático.....	31

Lista de siglas y abreviaturas

- UNSM-T** : Universidad Nacional de San Martín - Tarapoto.
- FISI** : Facultad de Ingeniería de Sistemas e Informática.
- IIAP** : Instituto de Investigaciones de la Amazonia Peruana.
- SI** : Sistema Informático
- INIA** : Instituto Nacional de Innovación Agraria.

Resumen

El presente proyecto de investigación titulado “Buscador Informático basado en el Modelo Vectorial para la Difusión de la Información Científica de las Especies de Palmeras (Palmae) Inventariadas de la Región San Martín”, fue desarrollado con financiamiento de la UNSM-T el año 2017, con el objetivo de mejorar la difusión de la información científica de las especies de Palmeras inventariadas de la Región San Martín. Donde el tipo de investigación es aplicada – transversal, a nivel experimental, y de subnivel pre-experimental, de diseño pre y post test, para una muestra de 22 profesionales, utilizando el Modelo de Recuperación Vectorial para la construcción del Buscador Informático (www.palmerassanmartin.com).

Palabras clave: Buscador Informático, Modelo Vectorial, Difusión de la Información, Inventario de Especies de Palmeras.

Abstract

The following research project titled as "Computer Seeker based on the Vector Model for the Dissemination of Scientific Information on the Species of Palms (Palmae) Inventoried in the San Martin Region", was developed with funding from the UNSM-T in 2017, with the objective of improving the dissemination of scientific information of the inventoried palm species of the San Martín Region. The following is an applied - transversal, at experimental level research, and pre-experimental sub-level with a pre and posttest design, for a sample of 22 professionals, using the Vector Recovery Model for the construction of the Computer Search Engine (www.palmerassanmartin.com).

Keywords: Computer Searcher, Vector Model, Information Dissemination, Inventory of Palm Species.



Introducción

El Proyecto de Investigación titulado “Buscador Informático basado en el Modelo Vectorial para la difusión de Información Científica de las Especies de Palmeras (Palmae) Inventariadas de la Región San Martín”, presentado por la autora Bach. Carolina Mendoza Amasifuen en el marco del Reglamento de concurso de Proyectos de Investigación a Nivel de Pre Grado, financiados por la UNSM-T período 2017, tuvo como propósito mejorar la difusión de la información científica de las especies de Palmeras inventariadas de la Región San Martín, se desarrolló entre los meses de julio a diciembre del año 2017, se inició conociendo las bases de datos de información científica de las especies de Palmeras Inventariadas de la Región San Martín con el apoyo del Instituto de Investigaciones para la Amazonia Peruana (IIAP), luego se implementó el modelo de recuperación vectorial basado en el cálculo de la fórmula de similaridad entre las consultas y los vectores de información de cada especie, y finalmente analizamos la influencia del buscador informático en la difusión de información científica de las especies de Palmeras Inventariadas de la Región San Martín.

El Estudio fue de tipo Aplicada-Transversal, a Nivel Experimental, porque se trabajó con un solo grupo de 22 profesionales especializados denominado experimental, correspondiente a la muestra.

Según el Ministerio del Ambiente en el año 2012, el Perú es uno de los países con mayor diversidad de ecosistemas y especies de flora del mundo, distribuidos en una gran diversidad de geoformas, suelos y climas. Destacan los bosques amazónicos, con aproximadamente 70 millones de hectáreas; los herbazales altoandinos, con 19 millones de hectáreas, y los matorrales andinos, con 9 millones de hectáreas, no obstante, la difusión de la información en taxonomía, distribución y ecología de las especies, son escasos, por lo que los recursos biológicos de la zona se encuentran insuficientemente valorados. Por tal motivo este proyecto de investigación se orientó a crear una herramienta tecnológica que ayude a mejorar el proceso de difusión, poniendo énfasis en investigar cómo implementaríamos un buscador informático basado en el modelo de recuperación vectorial que se ajuste para tal fin. Los resultados que obtuvimos en nuestro proyecto demuestran que se ha logrado mejorar la difusión de la información científica de las especies de palmeras inventariadas en nuestra región.

El informe final de este proyecto de investigación, consta de tres (3) capítulos, el primero referido a la Revisión Bibliográfica, el segundo a los Materiales y Métodos, y el tercero a los Resultados de la Investigación.

CAPÍTULO I

REVISIÓN BIBLIOGRÁFICA

1.1. Fundamento teórico científico

1.1.1. Buscador informático

Son los programas dentro de un sitio o página web, los cuales, al ingresar palabras clave, operan dentro de la base de datos del mismo buscador y recopilando todas las páginas posibles, que contengan información relacionada con lo que se esté buscando.

Los buscadores dependen de un programa que se encarga de hacer coincidir las palabras clave que el usuario indica en la búsqueda con documentos más relevantes que existen en la red.

Tipos de buscadores

Los buscadores se pueden clasificar en tres tipos, según la forma de obtener las direcciones que almacenan en su base de datos.

- **Índices de búsqueda:** En los índices de búsqueda, la base de datos con direcciones la construye un equipo humano. Es decir, un grupo de personas va rastreando la red en busca de páginas. Vistas éstas son clasificadas por categorías ó temas y subcategorías en función de su contenido. De este modo, la base de datos de un índice de búsqueda contiene una lista de categorías y subcategorías relacionadas con un conjunto de direcciones de páginas web que tratan esos temas.
- **Motores de búsqueda:** En este caso, el rastreo de la web lo hace un programa, llamado *araña* ó motor (de ahí viene el nombre del tipo de buscador). Este programa va visitando las páginas y, a la vez, creando una base de datos en la que relaciona la dirección de la página con las 100 primeras palabras que aparecen en ella. Como era de esperar, el acceso a esta base de datos se hace por palabras clave: la página del buscador me ofrece un espacio para que yo escriba la o las palabras relacionadas con el tema que me interesa, y como resultado me devuelve directamente un listado de páginas que contienen esas palabras clave.

- **Metabuscadores:** Los metabuscadores son páginas web en las que se nos ofrece una búsqueda sin que haya una base de datos propia detrás: utilizan las bases de varios buscadores ajenos para ofrecernos los resultados. Un ejemplo de metabuscador es Metacrawler.

1.1.2. Modelos para la recuperación de información

Bajo el término modelos de recuperación se engloban todas aquellas técnicas que tienen por objetivo facilitar el proceso de comparación entre una consulta determinada y un conjunto de textos sobre los que se realiza la consulta. Se trata de herramientas muy utilizadas dentro del campo de la recuperación y organización de la información.

1.1.2.1. Modelo Booleano

Es un modelo de recuperación simple, basado en la teoría de conjuntos y el álgebra booleana. Dada su inherente simplicidad y su pulcro formalismo ha recibido gran atención y sido adoptado por muchos de los primeros sistemas bibliográficos comerciales. Su estrategia de recuperación está basada en un criterio de decisión binario (pertinente o no pertinente) sin ninguna noción de escala de medida, sin noción de un emparejamiento parcial en las condiciones de la pregunta.

1.1.2.2. Modelo Vectorial

El modelo de recuperación vectorial o de espacio vectorial propone un marco en el que es posible el emparejamiento parcial, asignando pesos no binarios a los términos índice de las preguntas y de los documentos. Estos pesos de los términos se usan para computar el grado de similitud entre cada documento guardado en el sistema y la pregunta del usuario.

En este modelo de recuperación de información, cada documento se representa a través de un vector de n dimensiones cuyas componentes son los términos que aparecen en el texto. El valor de cada componente se calcula a partir del IDF (Inverse Document Frequency) y se obtiene una representación vectorial para las consultas, que se comparan con los vectores de los documentos empleando una función de similitud.

$$\text{sim}(d, q) = \cos \theta = \frac{\sum_{i=1}^t w(i, d) \times w(i, q)}{\sqrt{\sum (w(i, d))^2} \times \sqrt{\sum (w(i, q))^2}}$$

Norma de d
Norma de q

Figura 1: Fórmula del Modelo Vectorial (Fuente: Robertson, S. 2004)

1.1.2.3. Modelo Probabilístico

El modelo de recuperación probabilístico se basa en la equiparación probabilística, dados un documento y una pregunta, es posible calcular la probabilidad de que ese documento sea relevante para esa pregunta.

Se fundamenta en la representación binaria de los documentos, al igual que en el modelo de recuperación booleano, indicando presencia o ausencia de términos mediante 0 y 1. Su diferencia radica en el método estadístico y en las premisas bajo las que se constituye su funcionamiento estableciendo las siguientes aseveraciones:

Según la consulta planteada por el usuario, los documentos de la colección se clasifican en dos grupos; 1) Conjunto de Documentos Relevantes y 2) Conjunto de Documentos Irrelevantes.

Existe una respuesta ideal del sistema, constituida por el conjunto de documentos relevantes, a la que se denomina Conjunto de Respuesta Ideal.

Existe una Consulta Ideal, que es aquella que proporciona un Conjunto de Respuesta Ideal o lo que es lo mismo el conjunto de documentos relevantes para el usuario.

Aunque a priori se desconoce cuál es la Consulta Ideal (el usuario no tiene por qué conocerla), sí se sabe que es una combinación de 0 y 1 por ser un modelo binario de recuperación. Se desconocen por tanto los términos que se deberían introducir para obtener el Conjunto de Respuesta Ideal.

1.1.3. Modelo de recuperación vectorial

El modelo de recuperación vectorial o de espacio vectorial propone un marco en el que es posible el emparejamiento parcial a diferencia del modelo de

recuperación booleano, asignando pesos no binarios a los términos índice de las preguntas y de los documentos. Estos pesos de los términos se usan para computar el grado de similitud entre cada documento guardado en el sistema y la pregunta del usuario.

El modelo de recuperación vectorial fue presentado por Saltón en 1975 y posteriormente asentado en 1983 junto con Mc Gill y se basa en tres principios esenciales.

- La equiparación parcial, esto es, la capacidad del sistema para ordenar los resultados de una búsqueda, 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 adscribiendo a cada término en cada documento un número real que refleje su importancia 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 reflejen la importancia de los mismos en relación a su necesidad informativa.

Factor TF: Term Frequency: Frecuencia de Aparición de un Término. El factor TF es la suma de todas las ocurrencias o el número de veces que aparece un término en un documento. A este tipo de frecuencia de aparición también se la denomina “Frecuencia de aparición relativa” por qué atañe a un documento en concreto y no a toda la colección. Su cálculo se efectúa una vez el texto del documento ha sido normalizado, según los procesos de depuración mencionados. Posteriormente se lleva a cabo el conteo de las veces que el término aparece presente en el documento. De hecho, se deja bien claro que es necesario calcular el TF de cada término en cada documento.

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

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

Figura 2: Cálculo del TF (Fuente: Robertson, S. 2004)

Factor IDF: Inverse Document Frequency: Frecuencia Inversa del Documento para un término. El factor IDF de un término es inversamente proporcional al número de documentos en los que aparece dicho término. Esto significa que cuanto menor 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 discriminatorio.

El Factor IDF es único para cada término de la colección. El IDF de un término dado (n) se realiza 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 en término (n) dado). Al valor resultante se le suma 1 para corregir los valores para los términos con IDF muy bajos (Aunque esta variación depende del sistema de recuperación).

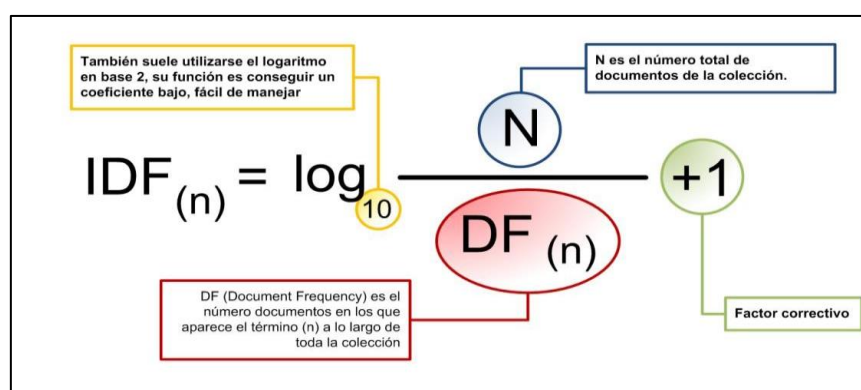


Figura 3: Fórmula del factor IDF (Fuente: Robertson, S. 2004)

Medida de la precisión media (MAP). Es una medida que intenta sumar todos los valores de los promedios de las precisiones, o sea que tan preciso es el modelo de recuperación de información web para retornar páginas web (documentos) que sean relevantes o importantes para el usuario. Su fórmula está dada por (Souza, R. R. 2006):

$$MAP = \frac{\sum_{n=1}^N P@n \times rel(n)}{r_q}$$

Figura 4: Fórmula MAP (Fuente: Robertson, S. 2004)

1.1.4. Difusión de información

La difusión implica propagar algo, una información, dato o noticia, con la misión de hacerlo público y de ese modo ponerlo en conocimiento de una importante cantidad de individuos que lo desconocen hasta ese momento.

No hay una forma única de difusión, sino diferentes tipos de productos y servicios capaces de vehicular la información hacia los usuarios. Con objeto de sistematizar se pueden distinguir dos formas básicas de difusión, la difusión bajo demanda y la difusión documental.

En cuanto a las vías de difusión puede ser: papel impreso, tabloneros de anuncios, expositores, soporte magnético para consultar en ordenadores, página web, correo electrónico, difusión verbal (persona a persona, conferencias, cursos), medios audiovisuales (videos informativos).

1.1.4.1. La difusión bajo demanda

Se trata de una difusión pasiva. La iniciativa parte del usuario que necesita información. Se realiza cuando el usuario se dirige al centro con el objetivo de solicitar una información concreta. El documentalista interroga al sistema y recupera la información solicitada. Las búsquedas bibliográficas y documentales responden a esta necesidad de información planteada por el usuario. Pueden realizarse sobre base de datos documentales creadas por el propio centro, pero, también, sobre bases externas.

Se considera también difusión bajo demanda la consulta en línea a una base de datos por parte del usuario sin la mediación de un documentalista.

1.1.4.2. La difusión documental

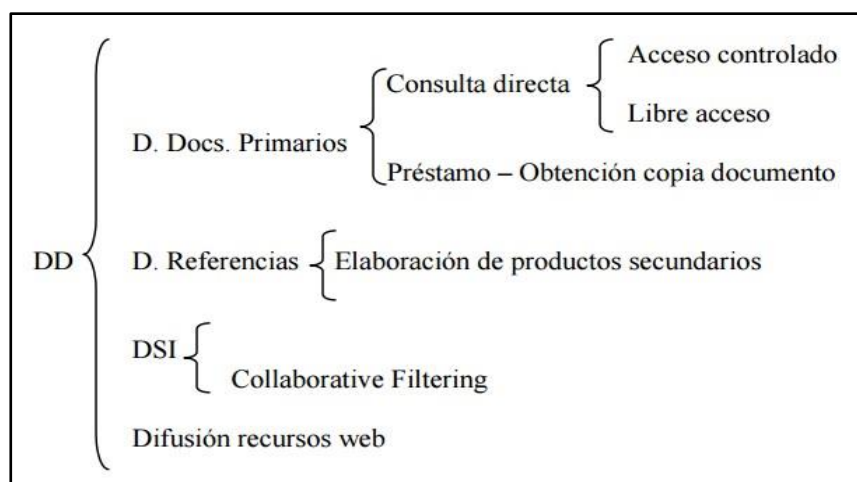


Figura 5: Difusión Documental (Fuente: www.uv.es/macass/T6.pdf)

La difusión documental es una difusión activa. Es aquella que implica iniciativa por parte del centro ofreciéndole, a los usuarios, productos documentales que juzga útiles, según el análisis que se haya hecho de sus necesidades, o señalándole sus problemas de información y ayudándole a resolverlos.

El centro deberá estudiar a sus usuarios con el objeto de averiguar qué tipo de información necesitan, en conjunto o por grupos, y debe estudiar también el sistema de difusión más adecuado para cada necesidad.

Dentro de la difusión documental se puede distinguir entre difusión de documentos primarios, la difusión de documentos secundarios o de referencia de documentos y la difusión selectiva de la información.

1.1.4.3. Difusión de documentos primarios

Consulta directa en el centro: En acceso controlado. El usuario puede consultar los ficheros, pero debe solicitar el documento que necesite a un empleado del centro. Se aplica a colecciones de documentos especiales por su contenido o soporte, obras raras o únicas, materiales débiles, materiales valiosos.

Libre acceso. Los documentos se encuentran a disposición del usuario ordenados generalmente según criterios temáticos. Da más libertad al usuario, pero disminuye el control de los fondos y requiere más espacio.

También supone hacer frente a la complejidad que supone la ordenación de los fondos.

Préstamos: El usuario obtiene el documento y puede utilizarlo durante un período de tiempo determinado. El documento debe ser devuelto.

Obtención de copia del documento. El usuario obtiene una copia del documento que pasa a ser de su propiedad, no debe devolverlo.

1.1.4.4. Difusión de referencia de documentos

La mayoría de centros de documentación elabora algún tipo de publicación secundaria que contiene referencias de documentos y transmite a los usuarios para su conocimiento.

1.1.4.5. Difusión selectiva de la información

Es un sistema de difusión “a la carta”, por el cual se ofrece a cada usuario las referencias de documentos correspondientes a sus temas de interés seleccionados a partir de todos los documentos recibidos durante un determinado período.

Proporciona a los usuarios una información personalizada, periódica y selectiva, sobre todo lo nuevo que se vaya publicando en temas concretos. Cada vez que se incorporan documentos nuevos que responden al interés del usuario se les envía. Para ello es necesario realizar el denominado perfil de interés de cada usuario. El usuario recibe referencias sin necesidad de demandar o acudir al centro cada vez.

Puede ser individual o colectiva: la individual responde al perfil de un solo usuario y la colectiva a un grupo homogéneo de usuarios.

La difusión selectiva de información (DSI) ha sido tradicionalmente un problema no resuelto de Centros de Documentación, Centros de Información y Bibliotecas. Los perfiles de usuarios están, la información está, pero mantener un servicio de diseminación selectiva funcionando en forma eficiente consume significativos recursos organizacionales y es casi inevitable que en un procesamiento manual o semiautomático, que implica la administración de listas de usuarios, la elaboración de índices y boletines, la producción y compaginación de fotocopias, el mantenimiento de listas de

direcciones de correo y teléfonos, la elaboración y etiquetados de sobres y el envío por correo o por fax, se cometan errores. No se puede pensar en atender requerimientos demasiado específicos con herramientas manuales o semiautomáticas.

Para realizar un DSI automatizado se requiere algún programa informático que integre en un único concepto el acceso y la búsqueda en la base de datos, la agenda del sistema operativo y el correo electrónico. En los DSI automatizados las consultas a las bases de datos se realizan automáticamente y las respuestas que se obtienen se envían a cada usuario registrado en la lista de correos asociada al servicio. El mensaje de correo tendrá en forma anexa un documento con las referencias de aquellos documentos que se ajusten al perfil de información que ha solicitado o bien los textos íntegros de los documentos si el centro posee una base de datos fuente textuales. Los servicios de difusión pueden ser diarios, semanales, mensuales, semestrales, anuales, etc.

1.1.4.6. Difusión de recursos Web

Difusión de recursos en internet. Consiste en seleccionar, ordenar y describir las webs de interés para los usuarios del centro. Se distribuye mediante las redes internas del centro o intranet. (Ahora se empieza a aplicar el término de Arquitectura de la información a esta actividad). También puede consistir en enviar, por correo electrónico, las URL de nuevas webs de interés (Pérez Álvarez, 1998)

1.1.5. Palmeras

Las palmeras son árboles que se hallan en zonas subtropicales o intertropicales que presentan frutos más bien carnosos, flores de color blanco y llegan a medir 30 metros de altura.

Las palmeras pertenecen al orden de las angiospermas (plantas que tienen el óvulo protegido en el interior del ovario). Dentro del orden de las angiospermas hay dos grupos diferenciados, las monocotiledóneas y las dicotiledóneas. Dentro del grupo de plantas monocotiledóneas encontramos a las palmeras, aparte de otras plantas como las gramíneas, las bromelias, los agaves, las

cordilinas, las yucas, etc. Plantas que por el aspecto similar al de las palmeras en ocasiones hay gente que las incluyen dentro de su mismo grupo erróneamente. (Montosa, 2015).

Como hemos dicho, las palmeras son monocotiledóneas, de la familia de las Arecaceae (antes llamadas Palmae) que se incluyen dentro del orden de las Arecales. Actualmente se estima que hay entre 2300 y 2500 especies dentro de unos 189 géneros, pero esto va cambiando a medida que se van descubriendo nuevas especies, o a través de la investigación de algunos científicos que demuestran mediante el ADN que algunas especies que estaban incluidas en otros géneros, no pertenecían a estos.

1.1.5.1. Morfología de las palmeras

A nivel morfológico dividimos la palmera en tres partes:

- Sistema radicular.
- Estípite o tronco.
- Copa o corona.

- **Sistema radicular**

Las raíces de las palmeras nacen de la base del tronco o bulbo basal. Como en el resto de las plantas las raíces ejercen las funciones esenciales de anclaje y absorción de agua y nutrientes.

Las palmeras muestran dos sistemas diferentes de raíces, el primero es el sistema radicular principal o seminal que se manifiesta en la fase inicial de la plántula, aunque suele durar poco tiempo, luego es sustituido por el sistema radicular secundario, este es mucho más importante para la palmera y se conoce por el nombre de Sistema radicular adventicio. Este sistema nace en la base del tronco, muy cerca de la superficie del suelo, aunque en algunas especies las raíces adventicias pueden verse por encima del suelo.

Este tipo de raíz es característico de las monocotiledóneas, forman un abundante haz en forma de cabellera de raíces finas y largas, pueden cubrir distancias de más de 10 metros respecto al tronco y algunas hasta 50 metros

de longitud, pudiendo alcanzar una profundidad de unos 5 metros, aunque normalmente no pasan de 2 metros.

- **Estípite o tronco**

A los troncos de las palmeras se le llama, estípite o estipe, este término en un ápice de crecimiento llamado yema apical. Este órgano es vital, ya que, si sufre algún daño o alteración fisiológica y muere, fallece la planta. Esto no pasa con las palmeras multicaules ya que cuando muere una de sus yemas apicales los hijuelos que permanecen en fase estacionaria empiezan a desarrollarse. El estípite de las palmeras, en lugar de tener corteza como los árboles tiene una epidermis endurecida que forma una capa protectora. La mayoría de las especies alcanzan el diámetro máximo del tronco antes de empezar a crecer en altura. Puede haber engrosamientos posteriores por el hinchamiento de las células o el robustecimiento de sus paredes.

Los troncos de las palmeras son muy fuertes, gracias a un anillo periférico de haces frondosos lignificados y tiene mucha flexibilidad debido al córtex blando de su interior.

La mayoría de las palmeras tienen un tronco de anchura uniforme y recto, aunque hay algunos que se estrechan hacia arriba, otros que son estrechos y parecen cañas de bambú, otros en forma de botella, etc.

La superficie del tronco se puede ver de diferentes formas, que pueden ser, lisas, cubiertas de fibras entrelazadas, de vainas persistentes de las hojas, con espinas, etc.

El punto del tronco donde se inserta la hoja se llama nudo, y el espacio entremedias internudo. Las cicatrices dejadas por las hojas caídas se denominan anillos. En las palmeras se puede calcular el vigor de crecimiento dependiendo de la distancia entre los anillos y pueden servir para calcular la edad que tiene. Cuando se observan los internudos cortos o anillos muy juntos, quiere decir que ha tenido un crecimiento lento, y cuando sucede lo contrario y los anillos están muy separados, significa que ha tenido un crecimiento rápido.

Se pueden diferenciar grupos de palmeras en función de la forma del tronco:

1. Unicaules: Palmeras con un único tronco, forman un grupo muy común y que incluye muchas especies.
2. Multicaules: Palmeras que producen hijuelos a nivel del suelo o por debajo de él, consiguiendo que surjan varios troncos.
3. Ramificadas: La ramificación epigea natural en las palmeras es inusual, reducida a solo siete géneros. Hay dos tipos de ramificación: por bifurcación del ápice de crecimiento (ramificación dicotómica, es decir, que se divide en dos partes) o por ramificación lateral.
4. Ramificación dicotómica: Se produce cuando el ápice de crecimiento se divide en dos partes iguales y surgen dos ramas. Este tipo de ramificación puede ser epigea (es decir, que se desarrolla sobre el suelo) o darse en troncos subterráneos.
5. Ramificación lateral: Se produce en especies con troncos subterráneos, rara vez se da en la parte aérea de los troncos de las palmeras, aunque puede verse ocasionalmente.
6. Acaules: Palmeras que carecen de tronco o lo tienen muy reducido. Suelen tener tallo, aunque no es visible por ser subterráneo. Este grupo suele ser de palmeras pequeñas.

• Copa

La copa de las palmeras la forman tres elementos, las hojas, el sistema o aparato reproductor y la yema apical.

1.2. Definición de términos básicos

1.2.1. Especie

Linaje (o conjunto de ellos) que ocupa una zona adaptativa mínimamente diferenciada de la de cualquier otro linaje de su rango, y que evoluciona separadamente de todos los linajes externos a su rango.

1.2.2. Información científica

El conjunto de los registros o resultados de la ciencia, del conocimiento basado en el método científico, grabados y dados a conocer por procedimientos digitales, en la red, pero también, incluso, por medios impresos en papel.

1.2.3. Inventario

Reconocimiento, ordenamiento, catalogación, cuantificación y mapeo de entidades naturales como genes, individuos, especies, poblaciones, comunidades, ecosistemas o paisajes.

CAPÍTULO II

MATERIALES Y MÉTODOS

2.1. Sistema de hipótesis

Hipótesis alterna:

(Ha): “Con el uso del Buscador Informático basado en el Modelo Vectorial Sí mejorará la difusión de información científica de las especies de Palmeras inventariadas de la Región San Martín”.

Hipótesis nula:

(Ho): “Con el uso del Buscador Informático basado en el Modelo Vectorial NO mejorará la difusión de información científica de las especies de Palmeras inventariadas de la Región San Martín”.

2.2. Sistema de variables

Se considera las siguientes variables de estudio:

2.2.1. Variable dependiente Y:

Difusión de información.

Indicadores:

Y1: Nivel de difusión

Unidad de medida: cuantitativa

Y2: Vía de difusión.

Unidad de medida: ordinal

Y3: Forma de difusión.

Unidad de medida: ordinal.

2.2.2. Variable independiente

X: Buscador Informático basado en el modelo Vectorial.

Indicadores:

X1: Similitud Consulta-Respuesta.

Unidad de medida: Porcentaje.

X2: Satisfacción del usuario

Unidad de medida: Cualitativa

Operacionalización de variables

Tabla 1

Escala de medición de las variables.

Variable	Indicador	Escala de medición
Dependiente:	Y1: Nivel de difusión	Cuantitativa discreta
Difusión de	Y2: Vía de difusión	Cualitativa ordinal
información	Y3: Forma de difusión	Cualitativa ordinal
Independiente:	X1: Similitud	Cuantitativa de razón
Buscador	Consulta-Respuesta	
Informático basado	X2: Satisfacción de	Cuantitativa de razón
en el Modelo	usuario	
Vectorial		

Fuente: Elaborado por el investigador

2.3. Tipo de método de la investigación

2.3.1 Tipo:

El estudio presentado corresponde a una investigación del tipo Aplicada - Transversal, porque los resultados obtenidos pretenden aplicarse en forma inmediata para ayudar a resolver la situación problemática planteada, analizando los resultados obtenidos en un mediano plazo.

2.3.2. Nivel:

Experimental. - Porque se controló la variable independiente con el fin de mejorar la difusión de la información científica de las especies de palmeras inventariadas de la Región San Martín con el uso de un Buscador Informático.

2.3.3. Subnivel:

Pre-Experimental. - Porque se trabajó con un solo grupo denominado experimental, correspondiente a la muestra, en donde se aplicó el Buscador Informático Basado en el Modelo Vectorial para la difusión de la información científica inventariada de la Región San Martín.

2.4. Diseño de investigación

El diseño de la investigación se basó en la realización de trabajo de medición pre y post; antes y después de la aplicación del Sistema de Información basado en el Modelo Vectorial.

Cuyo diagrama fue el siguiente:

G-----O1-----X-----O2

Dónde:

- G: Grupo Experimental. Evaluación pre test al grupo de estudio (Grupo experimental).
- O₁: Evaluación pre test: Aplicación o tratamiento (al grupo experimental).
- X: Buscador Informático basado en el modelo de recuperación Vectorial: Aplicación del Sistema Informático basado en el modelo vectorial.
- O₂: Evaluación pos test: Comparación de los resultados de la evaluación pre test y evaluación pos test.

2.5. Población y muestra

2.5.1. Población

Profesionales de las instituciones que solicitan información al Instituto de Investigación de la Amazonia Peruana (IIAP), con sede San Martín, específicamente en el Programa de investigación en Biodiversidad Amazónica (PIBA), quien es el encargado del estudio de las especies de Palmeras en la Región San Martín.

Existen 5 instituciones en la región San Martín que constantemente solicitan información al programa PIBA, éstas son: DRASAM, UNSM, IIAP, UPeU, INIA.

2.5.2. Muestra

La muestra fue entre 6 y 5 personas por cada institución que acuden a solicitar información de las especies de palmeras inventariadas de la región San Martín.

2.6. Técnicas e instrumentos de recolección de datos

Tabla 2

Técnicas e instrumentos a utilizar en el estudio.

Técnica	Justificación	Instrumentos	Aplicado en...
1. Encuesta	1. Permite conocer la difusión de información de las Palmeras	1. Cuestionario.	1. A la muestra que es parte del estudio.
2. Registros	2. Proporciona la información del buscador informático.	2. Buscador Informático	2. Procesos que se desarrollan dentro del buscador.
3. Análisis de documentos	3. Para obtener la información de las fuentes secundarias referentes a temas de la investigación.	3. Fichas bibliográficas. Subrayado.	3. La bibliografía necesaria para desarrollar el marco teórico y la información complementaria.

Fuente: Elaborado por el investigador.

2.7. Técnicas de procedimiento

Para hallar que hay en los datos:

La media aritmética o promedio (M): Es el estadístico de tendencia central más significativo y corresponde variables de cualquier nivel de medición, pero particularmente a las mediciones de intervalo y de razón.

$$M = \frac{X_1 + X_2 + X \dots}{N}$$

Dónde: M, media aritmética; X, frecuencia de un valor cualquiera de la variable y N, número total de los valores considerados.

Desviación Estándar. - Es una medida de dispersión de las variables cuantitativas.

Programa de computador. Se utilizó el programa SPSS Statistics 22, versión para Windows 8.

Tratamiento Estadístico. Los análisis estadísticos empleados fueron descriptivos e inferenciales, en la parte descriptiva se usó una tabla de frecuencia, gráficos y medidas como la media y desviación estándar. En la parte inferencial, para comparar los diferentes datos que se obtienen a través del instrumento elaborado para los fines de la investigación, también se usó la estadística T-student.

Presentación de datos

- **Tablas:** Para un mejor entendimiento de los datos se usó tablas distribuyéndolos en columnas y filas, agrupando los datos evitando duplicidades y superando el desorden. Servirán de ayuda visual que permitirán organizar los resultados de la investigación.
- **Figuras:** Importantes para expresar la tendencia de un hecho o fenómeno. Se usarán figuras estadísticas como barras, líneas, áreas, etc.

CAPÍTULO III

RESULTADOS Y DISCUSIÓN

3.1. Presentación de resultados

En nuestro estudio se llegó encuestar a 22 profesionales de las diferentes instituciones que hacen uso de la información de las especies de palmeras inventariadas de la región San Martín, llegándose a encontrar lo siguiente:

Tabla 3

Resumen de los indicadores de la difusión de Información, antes de la implementación del buscador informático.

Variable	Indicador	N	Porcentaje marginal
Vía de difusión	Malo	3	14%
	Regular	14	64%
	Bueno	5	23%
	Muy bueno	0	0%
	Excelente	0	0%
Forma de difusión	Malo	7	22.7%
	Regular	11	72.7%
	Bueno	4	4.5%
	Muy bueno	0	0.0%
	Excelente	0	0
Nivel difusión de información	Malo	2	9%
	Regular	17	77%
	Bueno	3	14%
	Muy bueno	0	0%
	Excelente	0	0%

Fuente: Ficha de recolección de datos.

En cuanto a la vía de difusión, antes de la implementación del buscador informático, el 64% de los profesionales lo considera como regular. (Ver Tabla 3)

En cuanto a la forma de difusión, antes de la implementación del buscador informático, el 72.7% lo califica como regular. (Ver Tabla 3)

En cuanto al nivel de difusión de información, se encontró que el 77% de los profesionales lo calificaron como regular y un 9% como 'Malo', esto antes de la implementación del buscador informático. (Ver Tabla 3)

• **Sobre la Vía de Difusión (Pre Test)**

Tabla 4

Vía de Difusión Pre Test

Vía de difusión	Frecuencia	Porcentaje
Malo	3	14%
Regular	14	64%
Bueno	5	23%
Muy bueno	0	0%
Excelente	0	0%
Total	22	100%

Fuente: Elaborado por el investigador.

Se observa que antes del uso del Buscador Informático 3 profesionales califican como ‘Malo’ a la Vía de Difusión que utiliza el Instituto de Investigaciones de la Amazonia Peruana (IIAP) al momento de difundir algún tipo de información acerca de las palmeras inventariadas de la Región San Martín, mientras que 14 profesionales lo califican de ‘Regular’ y solo 5 como ‘Bueno’ (Tabla 4).

Complementariamente, se aprecia en altos porcentajes (64% y 23%) que los profesionales calificaron de ‘Bueno’, a la Vía de Difusión el IIAP. (Figura 6).

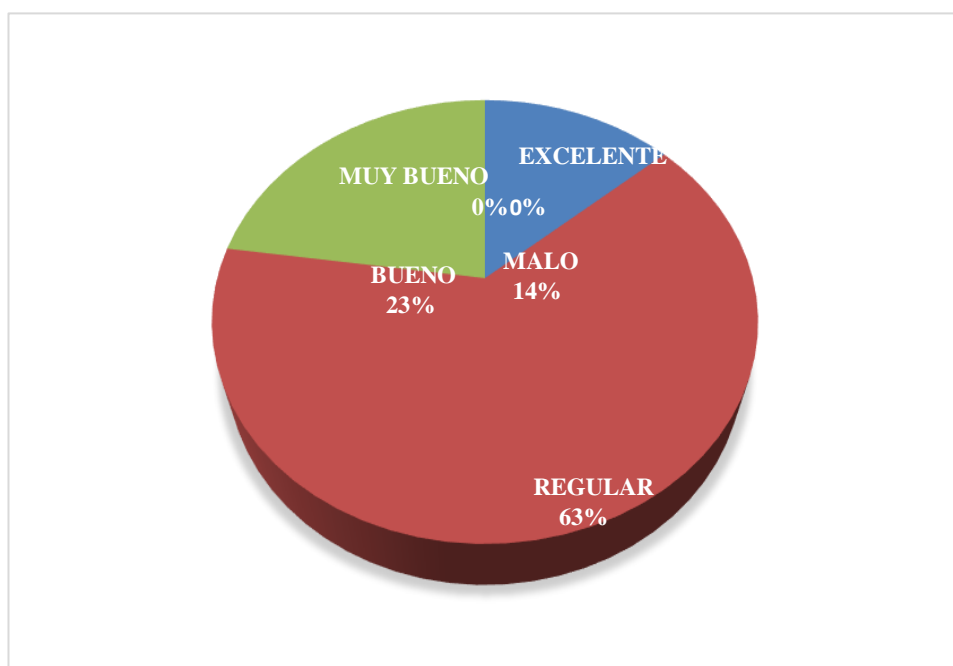


Figura 6: Vía de Difusión de la información Pre Test (Fuente: Tabla 4)

• **Sobre la Forma de Difusión (Pre Test)**

Tabla 5

Forma de Difusión Pre Test

Forma de difusión	Frecuencia	Porcentaje
Malo	7	31,8%
Regular	11	50,0%
Bueno	4	18,2%
Muy bueno	0	0%
Excelente	0	0%
Total	22	100%

Fuente: Elaborado por el investigador.

Se observa que antes del uso del Buscador Informático 7 profesionales califican como ‘Malo’ a la Forma de Difusión que utiliza el Instituto de Investigaciones de la Amazonia Peruana al momento de difundir algún tipo de información acerca de las palmeras inventariadas en la Región San Martín, mientras que 11 profesionales lo califican de ‘Regular’ y solo 4 como ‘Bueno’ (Tabla 5).

Complementariamente, se aprecia en altos porcentajes (31.8% y 50%) que los profesionales calificaron de ‘Malo’ y ‘Regular’, la forma de difusión que utiliza el buscador (Figura 7).

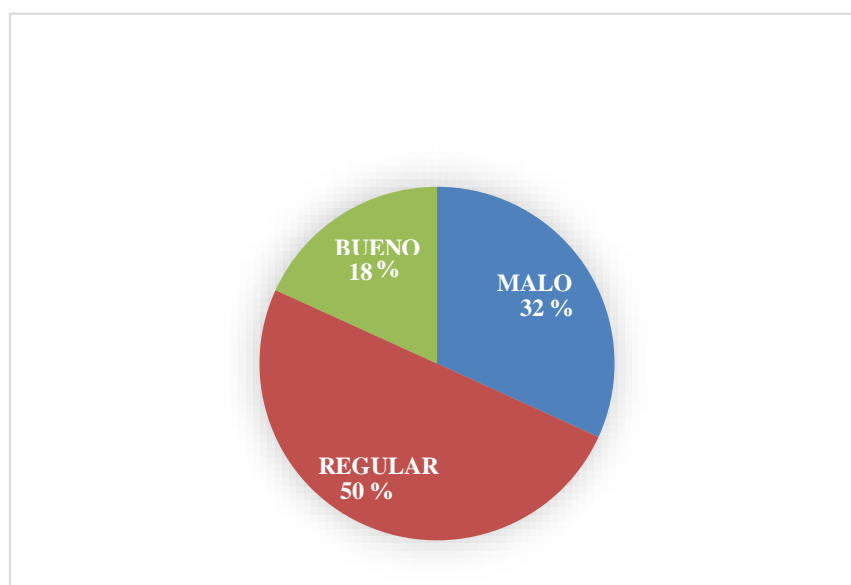


Figura 7: Forma de Difusión de la información Pre Test (Fuente: Tabla 5)

• **Sobre el Nivel de Difusión**

Tabla 6

Nivel de Difusión Pre Test

Nivel de difusión	Frecuencia	Porcentaje
Malo	2	9%
Regular	17	77%
Bueno	3	14%
Muy bueno	0	0%
Excelente	0	0%
Total	22	100%

Fuente: Elaborado por el investigador.

Se observa que antes del uso del Buscador Informático 2 profesionales califican como ‘Malo’ el Nivel de Difusión que alcanza el Instituto de Investigaciones de la Amazonia Peruana al momento de difundir algún tipo de información acerca de las palmeras inventariadas en la Región San Martín, mientras que 17 profesionales lo califican de ‘Regular’ y solo 3 como ‘Bueno’ (Tabla 6).

Complementariamente, se aprecia en altos porcentajes (77% y 14%) que los profesionales calificaron de ‘Malo’ y ‘Regular’, al Nivel de difusión que utiliza el buscador (Figura 8).

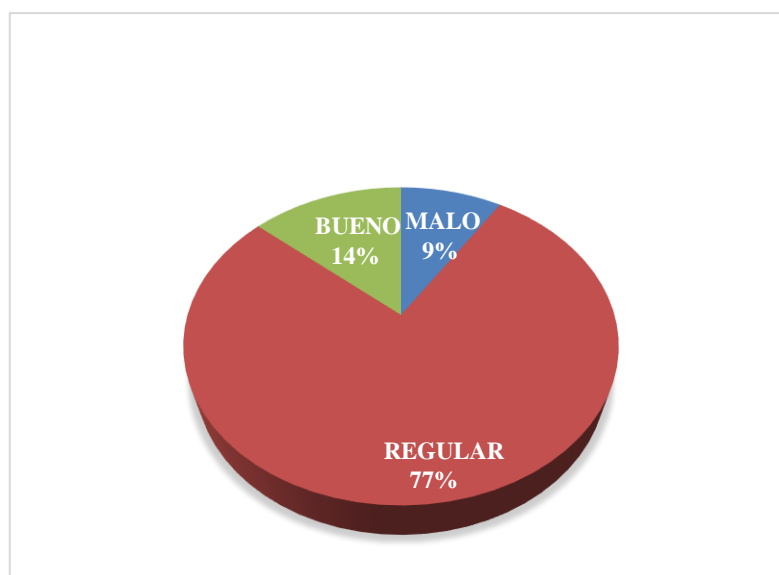


Figura 8: Nivel de Difusión de la información Pre Test (Fuente: Tabla 6)

Tabla 7

Resumen de los indicadores de la difusión de Información, después de la implementación de buscador informático

Variable	Indicador	N	Porcentaje marginal
Vía de difusión	Malo	0	0%
	Regular	0	0,0%
	Bueno	7	31,8%
	Muy bueno	12	54,5%
	Excelente	3	13,6%
Forma de difusión	Malo	0	0%
	Regular	1	4,5%
	Bueno	5	22,7%
	Muy bueno	11	50%
	Excelente	5	22,7%
Nivel difusión de información	Malo	0	0%
	Regular	0	0,0%
	Bueno	4	18,2%
	Muy bueno	13	59,1%
	Excelente	5	22,7%
Similitud consulta – respuesta	Malo	0	0,00%
	Regular	2	9,1%
	Bueno	10	45,5%
	Muy bueno	7	31,8%
	Excelente	3	13,6%
Satisfacción del usuario	Malo	0	0,0%
	Regular	0	0%
	Bueno	3	13,6%
	Muy bueno	17	77,3%
	Excelente	2	9%

Fuente: Ficha de recolección de datos.

Después de la implementación del buscador informático, el 54.5% de los profesionales calificaron a la vía de difusión como Muy bueno. (Ver Tabla 7)

Además, el 50% de los profesionales calificó como bueno a la forma de difusión utilizada después de la implementación del buscador informático. (Ver Tabla 7)

En cuanto al nivel de difusión de información, se encontró que el 59.1% de los profesionales lo calificaron como Muy bueno y el 22.7% como Excelente, esto después de la implementación del buscador informático. (Ver Tabla 7)

Con respecto al buscador informático, entre el 31.8% y el 13.6% de los profesionales calificó como bueno y excelente a la similitud entre lo requerido y los resultados obtenidos al momento de realizar una búsqueda. (Ver Tabla 7)

En cuanto a la satisfacción del usuario del buscador informático entre el 13.6% y el 77.3% de los profesionales lo calificó como ‘Bueno’ y ‘Muy bueno’. (Ver Tabla 7)

- **Sobre la Vía de Difusión**

Tabla 8
Vía de Difusión Post Test

Vía de difusión	Frecuencia	Porcentaje
Malo	0	0%
Regular	0	0,0%
Bueno	7	31,8%
Muy bueno	12	54,5%
Excelente	3	13,6%
Total	22	100%

Fuente: Elaborado por el investigador.

Se observa que después del uso del Buscador Informático 7 profesionales califican de ‘bueno’ a la vía de difusión que utiliza el buscador, mientras que 12 profesionales lo califican de ‘Muy bueno’ y 3 de ‘Excelente’ (Tabla 8).

Complementariamente, se aprecia en altos porcentajes (31.8% y 54.5%) que los profesionales calificaron de buena y muy buena, respectivamente, la vía de difusión que representa el buscador (Figura 9).

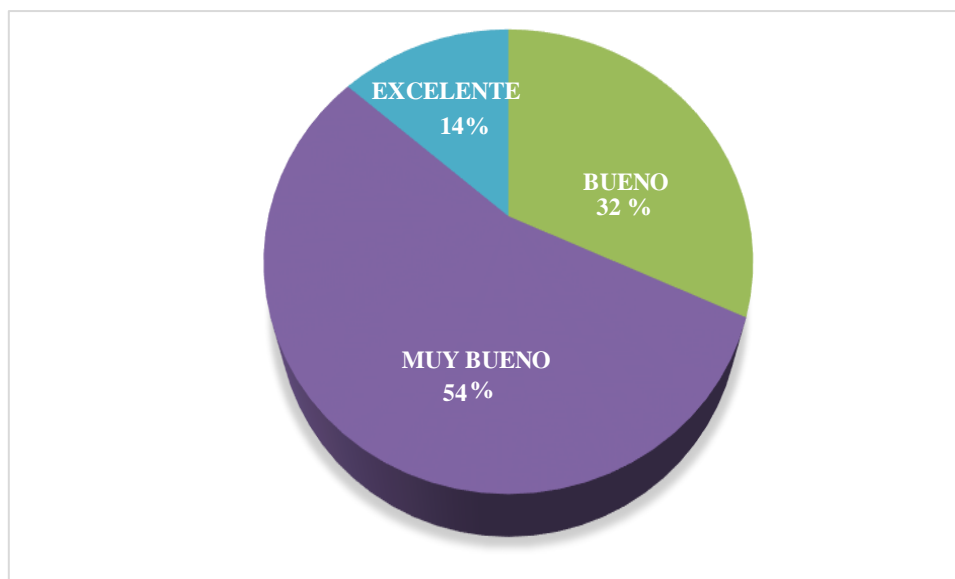


Figura 9: Vía de Difusión de la información Post Test (Fuente: Tabla 8)

- **Sobre la Forma de difusión**

Tabla 9

Forma de difusión Post Test

Forma de difusión	Frecuencia	Porcentajes
Malo	0	0%
Regular	1	4,5%
Bueno	5	22,7%
Muy bueno	11	50%
Excelente	5	22,7%
Total	22	100%

Fuente: Elaborado por el investigador

Se observa que después del uso del Buscador Informático 5 profesionales califican de 'bueno' a la forma de difusión que representa el sistema al momento de difundir algún tipo de información acerca de alguna palmera inventariada en la Región San Martín, mientras que 11 profesionales lo califican de 'Muy bueno' y 5 de 'Excelente' (Tabla 9).

Complementariamente, se aprecia en altos porcentajes (50% y 22.7%) que los profesionales calificaron de 'muy buena' y 'excelente', la forma de difusión que utiliza el buscador (Figura 10).

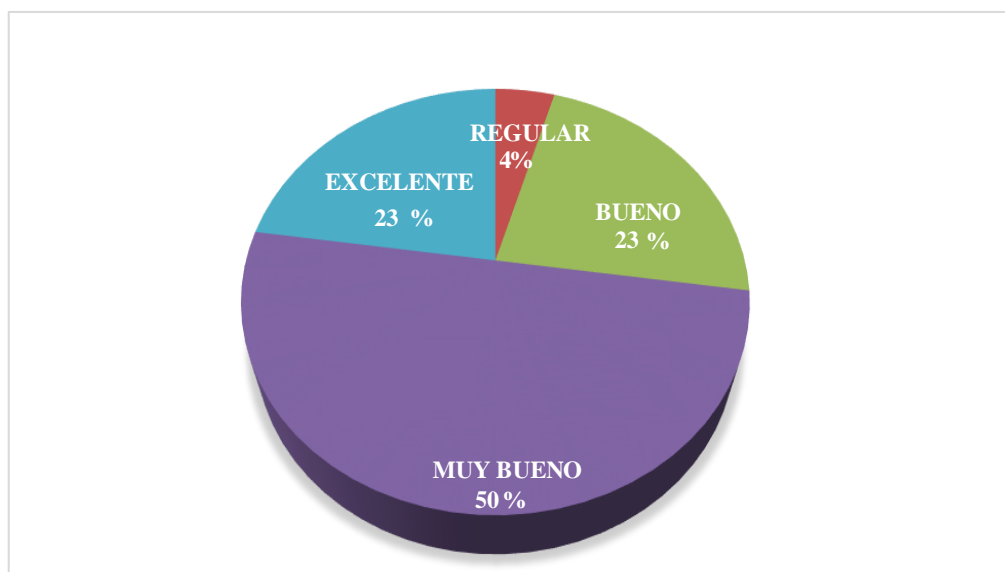


Figura 10: Forma de Difusión de la información Post Test (Fuente: Tabla 9)

• **Sobre el Nivel de difusión**

Tabla 10

Nivel de difusión Post Test

Nivel de difusión	Frecuencia	Porcentajes
Malo	0	0%
Regular	0	0,0%
Bueno	4	18,2%
Muy bueno	13	59,1%
Excelente	5	22,7%
Total	22	100%

Fuente: Elaborado por el investigador.

Se observa que después del uso del Buscador Informático 4 profesionales califican de 'bueno' al nivel de difusión que alcanza el sistema al momento de difundir algún tipo de información acerca de alguna palmera inventariada en la Región San Martín, mientras que 13 profesionales lo califican de 'Muy bueno' y 5 de 'Excelente' (Tabla 10).

Complementariamente, se aprecia en altos porcentajes (59.1% y 22.7%) que los profesionales calificaron de 'muy buena' y 'excelente', la forma de difusión que utiliza el buscador (Figura 11)



Figura 11: Nivel de Difusión de la información Post Test (Fuente: Tabla 10)

• **Sobre la Similitud Consulta-Respuesta (Post Test)**

Tabla 11

Nivel de difusión Post Test

Similitud Consulta- respuesta	Frecuencia	Porcentaje
Malo	0	0,00%
Regular	2	9,1%
Bueno	10	45,5%
Muy bueno	7	31,8%
Excelente	3	13,6%
Total	22	100%

Fuente: Elaborado por el investigador.

Se observa que después del uso del sistema informático solo 2 profesionales califican de 'Regular' a la similitud de los resultados brindados por el sistema al momento de buscar información sobre las palmeras inventariadas de la región San Martín, mientras que 10 profesionales lo califican de 'buena', 7 de 'muy buena' y 3 de 'excelente' (Tabla 11).

Complementariamente, se aprecia en altos porcentajes (45% y 31.8%) que los profesionales calificaron de 'Buena' y 'Muy buena', respectivamente, a la similitud de los resultados que brinda el sistema informático, en tanto que solo el 9.1%

consideró como regular los a la similitud de los resultados que brinda el buscador informático al momento de buscar información de las Palmeras inventariada de la región San Martín (Figura 12).

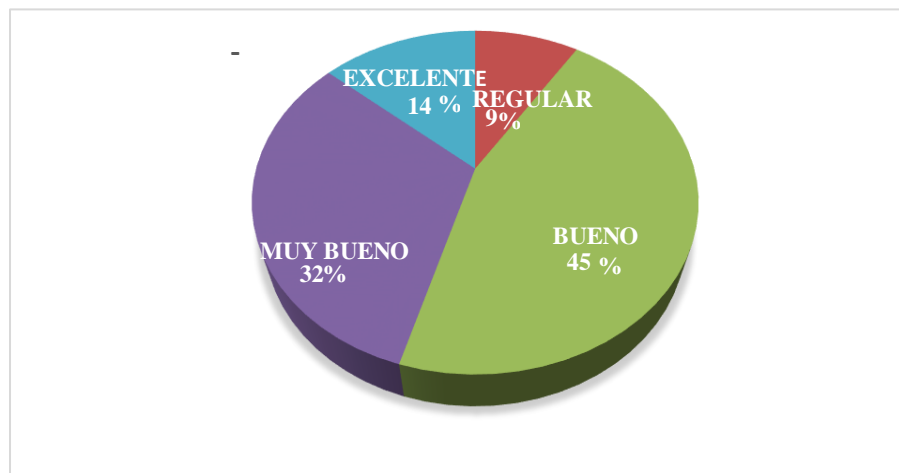


Figura 12: Similitud Consulta-Respuesta del Buscador Informático (Post Test (Fuente: Tabla 11)

• Sobre la Satisfacción del Usuario

Tabla 12

Sobre la Satisfacción del Usuario Post Test

Satisfacción del usuario	Frecuencia	Porcentaje
Malo	0	0,0%
Regular	0	0%
Bueno	3	13,6%
Muy bueno	17	77,3%
Excelente	2	9%
Total	22	100%

Fuente: Elaborado por el investigador.

Se observa que después del uso del buscador informático 3 profesionales califican de 'Bueno' a los resultados brindados por el sistema al momento de buscar información sobre las palmeras inventariadas de la región San Martín, mientras que 17 profesionales lo califican de 'Muy buena', 2 de 'Excelente' (Tabla 12).

Complementariamente, se aprecia en altos porcentajes (13% y 77.3%) que los profesionales calificaron de buena y muy buena, respectivamente, los resultados que

brinda el sistema informático, en tanto que solo el 9.1% consideró como regular los resultados que brinda el sistema informático al momento de buscar información de las palmeras inventariada de la región San Martín (Figura 13).

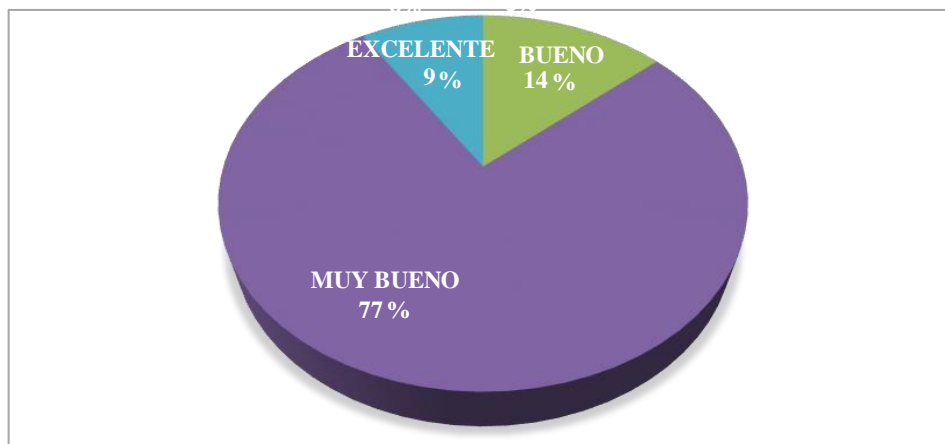


Figura 13: Satisfacción del Usuario Post Test (Fuente: Tabla 11)

Tabla 13

Comparación proporcional del nivel de difusión de información en relación con la implementación del buscador informático

Variable	Medición	Nivel de difusión (puntaje)	
Difusión de información	Pre test	Media	12.45
		n	22
	Pos test	Desv. Típ.	2.5
		Media	28.73
		n	22
		Desv. Típ.	4.13

Fuente: Ficha de recolección de datos

En relación con el nivel de difusión de información científica de las especies de palmeras inventariadas de la región San Martín se pudo constatar que mejoró en un 54% al incrementar su puntuación de 12.45 a 28.73 gracias a la implementación del buscador informático. (Ver Tabla13 y Figura 14)

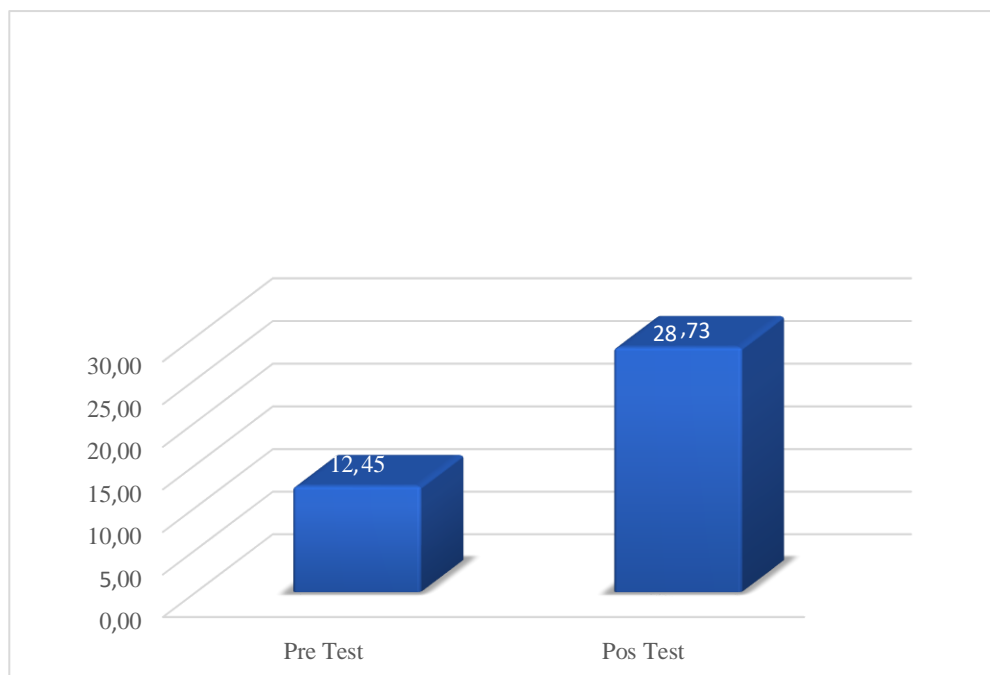


Figura 14: Nivel de difusión de información Pre y Post implementación del buscador informático (Fuente: Tabla 13)

Los datos son contundentes, la mejora en la difusión de información científica de las especies de palmeras inventariadas de la región San Martín es un claro reflejo de la implementación del buscador informático.

Por las razones antes presentadas, podemos afirmar que la hipótesis se confirma y por lo tanto se acepta, de modo que: “La implementación de un Buscador Informático basado en el Modelo de Recuperación Vectorial, mejoró la difusión de información científica de las especies de palmeras inventariadas de la Región San Martín”.

3.2. Contrastación de hipótesis

Hipótesis alterna:

(Ha): “Con el uso del Buscador Informático basado en el Modelo Vectorial Sí mejorará la difusión de información científica de las especies de Palmeras inventariadas de la Región San Martín”.

Hipótesis nula:

(H₀): “Con el uso del Buscador Informático basado en el Modelo Vectorial NO mejorará la difusión de información científica de las especies de Palmeras inventariadas de la Región San Martín”.

Por lo tanto:

$$H_0: \quad \mu_{pre} = \mu_{pos}$$

$$H_a: \quad \mu_{pre} \neq \mu_{pos}$$

3.2.1. Nivel de significancia

Se trabajó con un Nivel de significación del $|\alpha|=0.05$

3.2.2. Estadístico de prueba

Estadístico de Prueba:

$$\bullet \mu_0 = \frac{\overline{\chi_{pre}} - \overline{\chi_{pos}}}{\sqrt{S_{p^2} \left(\frac{1}{n_{pre}} + \frac{1}{n_{pos}} \right)}}$$

Donde:

μ_0 = valor que representa a todo Pre y Pos test

$\overline{\chi_{pre}}$ = media del Pre test (variable dependiente)

$\overline{\chi_{pos}}$ = media del Pos test (variable dependiente)

S_{p^2} = varianza de todo Pre y Pos Test

$$S_{p^2} = \frac{S^2_{pre} * n_{pre} + S^2_{pos} * n_{pos}}{n_{pre} + n_{pos} - 2}$$

Donde:

S^2_{pre} = varianza del pre

S^2_{pos} = varianza del pos

n_{pre} = total de encuestados del pre

n_{pos} = total de encuestados del pos

Entonces:

$$S_{p^2} = \frac{17.04 * 22 + 17.25 * 22}{22 + 22 - 2}$$

$$S_p^2 = 17.96$$

$$\mu_0 = \frac{13.23 - 28.73}{\sqrt{17.96\left(\frac{1}{22} + \frac{1}{22}\right)}}$$

$$\mu_0 = -12.12$$

Par ordenado:

$$t_{(n_{pre}+n_{pos}-2, 1-\alpha/2)}$$

3.2.3. Determinación de la región crítica:



Figura 15: Región Crítica Distribución T-Student (Fuente: Elaboración propia).

Regla de decisión:

- Aceptar H_0 si $\mu_0 \in$ a la *Region de Aceptación*.
- Rechazar H_0 si $\mu_0 \notin$ a la *Region de Aceptación*.
- Valor experimental

$$\mu_0 = -12.013$$

- Se decide rechazar H_0 porque no pertenece a la Zona de Aceptación y se acepta H_a .
- Conclusión: Se concluye que se acepta la H_a , por tanto, con el uso del Buscador Informático basado en el Modelo Vectorial SÍ se mejoró la difusión de información científica de las especies de Palmeras inventariadas de la Región San Martín.

3.3. Discusión de resultados

La presente investigación evaluó la difusión de la información científica de las especies de palmeras inventariadas de la Región San Martín con la implementación de un Buscador Informático basado en el Modelo de Recuperación Vectorial, encontrándose lo siguiente: El 77 % de los profesionales indicaron que los resultados que brinda el buscador informático son 'Muy buenos' y esto ha contribuido a mejorar la difusión de la información científica de las especies de palmeras inventariadas de la región San Martín. (Tabla 7).

En efecto el 31.8% de los profesionales afirmó como bueno y excelente a la similitud entre lo requerido y los resultados obtenidos al momento de realizar una búsqueda. (Tabla 7)

Por otra parte, el 51 % de los profesionales afirmó que el buscador ha contribuido de forma efectiva el nivel de difusión de la información sobre las especies de palmeras, esto ha representado un incremento significativo del 54%. (Tabla 13)

El buscador informático ha contribuido en gran medida además en el conocimiento de nuevas especies de palmeras, pues contiene una amplia base de datos con las características y los usos, la cual se puede ir ampliando aún más con el tiempo.

Con respecto al buscador informático, se demostró satisfactoriamente que brinda resultados excelentemente válidos y además con una excelente a la similitud entre lo requerido y los resultados obtenidos al momento de realizar una búsqueda. (Tabla 7).

Por todo lo anterior y de acuerdo a los resultados expuestos, se pudo verificar la hipótesis de la investigación, de modo que se comprobó que si se implementa un Buscador Informático basado en Modelo de Recuperación Vectorial se mejora la difusión de la información científica de las especies de palmeras inventariadas de la Región San Martín.

CONCLUSIONES

Después del análisis y discusión de los resultados obtenidos en el presente trabajo de investigación, llegamos a las siguientes conclusiones:

- La forma de difusión utilizado por el Instituto de Investigaciones de la Amazonia Peruana (IIAP), para la difusión de la información científica de las especies de palmeras inventariadas de la región San Martín, es deficiente debido a que lo realizan a través de revistas impresas, que en muchas ocasiones solo llega a un grupo selecto de la población y que no está disponible en forma permanente.
- Incorporar herramientas tecnológicas a medida de las necesidades de nuestra Región, genera un impacto positivo y en este caso mejoramos significativamente la difusión de la información científica de las especies de palmeras inventariadas de la región San Martín. Esto permite conocer más sobre esta familia de las Palmeras, las condiciones actuales de los recursos, los vacíos de información sobre ellos y las amenazas permanentes; incrementa sustancialmente los conocimientos sobre las Palmeras en diversas áreas de las ciencias, podemos tener una aproximación sobre la verdadera potencialidad de estos recursos.
- La implementación del buscador informático, ayuda a evitar la duplicidad de esfuerzos en la elaboración de inventarios, permitiendo de esa manera invertir en nuevos estudios, que nos permitan conocer es lo que tenemos, dónde está distribuido y en qué estado se encuentra, para definir qué se puede hacer y cómo es que debemos actuar.
- La aplicación del Modelo de Recuperación Vectorial para el buscador informático de difusión de la información científica de las especies de palmeras inventariadas de la Región San Martín, nos ha permitido determinar que éste, es un buen modelo a usar, debido a que se adecua rápidamente al proceso de implementación de un sistema experto en recuperación. (www.palmerassanmartin.com). a mayor información que se suministre, mayor es su poder de búsqueda, recomendando a los usuarios información relevante para cada uno de ellos.

RECOMENDACIONES

Finalmente se formulan las siguientes recomendaciones:

- Realizar investigaciones que incorporen las tecnologías de información y comunicación que ayuden a difundir la información científica de las especies de palmeras de la región San Martín.
- Investigar nuevos modelos computacionales de aprendizaje de máquina que permitan implementar sistemas expertos en recuperación.
- Crear nuevas herramientas tecnológicas que apoyen a la difusión de la información científica de las especies de palmeras.
- Utilizar el Modelo de Recuperación Vectorial, para la difusión de otros inventarios realizados a nivel nacional, tales como flora y fauna.
- El Gobierno Regional debe invertir en este tipo de herramientas tecnológicas, que permiten mejorar la difusión de la información.
- La Universidad Nacional de San Martín en conjunto con el Instituto de Investigaciones de la Amazonia Peruana, debe darle continuidad a este tipo de trabajos, incrementando la información de su base de datos, porque cuanto más información posea mayor será su poder de búsqueda.

REFERENCIAS BIBLIOGRÁFICAS

- Canhos, V. P., de Souza, S., De Giovanni, R., & Canhos, D. A. L. (2004). Global Biodiversity Informatics: setting the scene for a “new world” of ecological forecasting. *Biodiversity Informatics*, 1.
- Carmen Soguero Pamplona. (2015). Concepto de Buscador. 28/05/2017, de Aula 21 Sitio web: <http://www.aula21.net/tallerwq/buscadores/buscador1.htm>
- Castro Hidalgo, M., & Viale López, J. A. (Febrero de 2005). Sistema de búsqueda y comparación de documentos. Universidad Peruana de Ciencias Aplicadas.
- David Miguel Campos Ramírez. (2007). Recuperación y Organización de la Información Modelos de Recuperación I. 18/05/2017, de David Miguel Campos Ramírez Sitio web: <http://modelos-recuperacion.50webs.com/recuperacion-modelo-vectorial.html>
- Juan Ramón Montosa. (2015). Filotaxis en palmeras. 21/05/2017, de Juan Ramón Montosa Sitio web: <http://palmerasonline.blogspot.pe/>
- López Monroy, A. (Enero 2013). Recuperación de información para respuesta a preguntas en documentos legales. Instituto Politécnico nacional centro de investigación en computación
- Lupita. (2010). Buscador web. 18/05/2017, de Informática Sitio web: <http://www.informatica-01a1.blogspot.pe/2010/04/que-es-un-buscador.html>
- Martínez Méndez, F. J. (2002). Propuesta y desarrollo de un modelo para la evaluación de la recuperación de información en internet. Tesis Doctoral, Universidad de Murcia. España.
- Ministerio del Ambiente del Perú. (2012). Memoria descriptiva del Mapa de Cobertura Vegetal del Perú. Lima, Perú: MINAM.
- Pelta, D. A. (2013). Algoritmos heurísticos en bioinformática.
- Perfectti, F. (2002). Especiación: modos y mecanismos. Soler M., Evolución: La base de la biología. Proyecto Sur. España.
- Robertson, S. (2004). Understanding Inverse Document Frequency: On theoretical arguments for IDF. *Journal of Documentation*. Vol.60: (5), 503-520 pp.
- Saralegi Urizar, X., & Alegria Loinaz, I. (2007). Similitud entre documentos multilingües de carácter científico-técnico en un entorno Web. Universidad del País Vasco, 71-78.

- Souza, R. R. (2006). Sistemas de recuperación de información de mecanismos de búsqueda web: panorama atual y tendencias. *Perspectivas en Ciencia da Informação (Impreso)*, Belo Horizonte, 11(2), 161-173.
- Tong, S., & Chang, E. (2001, Octubre). Support vector machine active learning for image retrieval. In *Proceedings of the ninth ACM international conference on Multimedia* (pp. 107-118). ACM.
- Villareal, H. M., Álvarez, M., Córdoba-Córdoba, S., Escobar, F., Fagua, G., Gast, F., ... & Umaña, A. M. (2004). *Manual de métodos para el desarrollo de inventarios de biodiversidad*.

ANEXOS



Anexo A: Reunión de coordinación con el biólogo Ángel Martín Rodríguez del Castillo, responsable de PIBA.



Anexo B: Aplicación de los instrumentos



Anexo C: Presentación de Buscador Informático



ANEXO D

ENCUESTA PRE - TEST

PRE-TEST

Ítems para evaluar la difusión de información científica de las especies de Palmeras inventariadas de la región San Martín.

*Estimado encuestado, solicitamos su colaboración para el presente cuestionario. Marque con una X la letra que corresponda según su criterio a la opción que represente más su opinión.

Cuestionario dirigido a los investigadores de las instituciones que requieren la información sobre las especies de palmeras de la regio San Martín.

Institución.....

Fecha:/..../....

• SOBRE LA DIFUSIÓN DE INFORMACIÓN

▪ Sobre la vía de difusión

1. La posibilidad de obtener toda la información sobre las especies de palmeras inventariadas por el IIAP es:
a) Malo b) Regular c) Buena d) Muy Buena c) Excelente
2. El tiempo requerido para obtener la información sobre las especies de palmeras inventariadas por el IIAP es:
a) Malo b) Regular c) Buena d) Muy Buena c) Excelente
3. La disponibilidad de la información de las especies de palmeras inventariadas por el IIAP es:
Malo b) Regular c) Buena d) Muy Buena c) Excelente
4. La posibilidad de tener libre acceso a la información sobre las especies de palmeras inventariadas por el IIAP es:
a) Malo b) Regular c) Buena d) Muy Buena c) Excelente

▪ Sobre la forma de difusión



5. La información sobre las especies de palmeras inventariadas se encuentra organizada para su consulta de forma:
Malo b) Regular c) Buena d) Muy Buena c) Excelente
6. La posibilidad de buscar la información sobre las especies de palmeras utilizando diferentes criterios de búsqueda (nombre científico, nombre común, ubicación geográfica, etc.) es:
a) Malo b) Regular c) Buena d) Muy Buena c) Excelente
7. Con respecto a la modalidad (folletos, revistas, conferencias, etc.) que usa el IIAP para transmitir la información de las especies de palmeras. ¿Cómo calificaría usted a esta modalidad?
a) Malo b) Regular c) Buena d) Muy Buena c) Excelente
8. La utilidad de la información sobre especies de palmeras ofrecidos por el IIAP es:
a) Malo b) Regular c) Buena d) Muy Buena c) Excelente

Te lo Agradezco, ¡Muchas Gracias!



ANEXO E

ENCUESTA POS - TEST

Pos-Test

Ítems para evaluar la difusión de información científica de las especies de palmeras inventariadas de la región San Martín.

*Estimado encuestado, solicitamos su colaboración para el presente cuestionario. Marque con una X la letra que corresponda según su criterio a la opción que represente más su opinión.

Cuestionario dirigido a los investigadores de las instituciones que requieren la información sobre las especies de palmeras de la región San Martín.

Institución.....

Fecha.../.../....

• SOBRE LA DIFUSIÓN DE INFORMACIÓN

▪ Sobre la vía de difusión

1. La posibilidad de obtener toda la información sobre las especies de palmeras inventariadas por el IIAP es:
a) Malo b) Regular c) Buena d) Muy Buena c) Excelente
2. El tiempo requerido para obtener la información sobre las especies de palmeras inventariadas por el IIAP es:
a) Malo b) Regular c) Buena d) Muy Buena c) Excelente
3. La disponibilidad de la información de las especies de palmeras inventariadas por el IIAP es:
a) Malo b) Regular c) Buena d) Muy Buena c) Excelente
4. La posibilidad de tener libre acceso a la información sobre las especies de palmeras inventariadas por el IIAP es:
a) Malo b) Regular c) Buena d) Muy Buena c) Excelente

▪ Sobre la forma de difusión

5. La información sobre las especies de palmeras inventariadas se encuentra organizada para su consulta de forma:
a) Malo b) Regular c) Buena d) Muy Buena c) Excelente

6. La posibilidad de buscar la información sobre las especies de palmeras utilizando diferentes criterios de búsqueda (nombre científico, nombre común, ubicación geográfica, etc.) es:
- a) Malo b) Regular c) Buena d) Muy Buena e) Excelente
7. Con respecto a la modalidad (folletos, revistas, conferencias, etc.) que usa el IIAP para transmitir la información de las especies de palmeras. ¿Cómo calificaría usted a esta modalidad?
- a) Malo b) Regular c) Buena d) Muy Buena e) Excelente
8. La utilidad de la información sobre especies de palmeras ofrecidos por el IIAP es:
- a) Malo b) Regular c) Buena d) Muy Buena e) Excelente

• **SOBRE EL BUSCADOR INFORMÁTICO**

✦ **Sobre la Similitud Consulta-Respuesta**

9. ¿Cuál diría usted que es el nivel de similitud o coincidencia entre lo que usted requiere y los resultados obtenidos?
- a) Malo b) Regular c) Buena d) Muy Buena e) Excelente

✦ **Sobre la Satisfacción de usuario**

10. ¿Qué tan satisfecho se encuentra usted con el uso del Buscador Informático de las especies de Palmeras?
- a) Muy insatisfecho
b) Insatisfecho
c) Ni satisfecho ni insatisfecho
d) Satisfecho
e) Muy satisfecho

Te lo Agradezco, ¡Muchas Gracias!