

UNIVERSIDAD NACIONAL DE LA AMAZONÍA PERUANA

Escuela de Post Grado

Maestría en Ingeniería de Sistemas con mención en Gerencia de Tecnología de la Información y Gestión de Software



“EFICIENCIA EN LAS BÚSQUEDAS DE INFORMACIÓN SOBRE
BIODIVERSIDAD AMAZÓNICA CON EL USO DE UN BUSCADOR SEMÁNTICO
EN LOS ESTUDIANTES DE CIENCIAS BIOLÓGICAS DE LA UNIVERSIDAD
NACIONAL DE LA AMAZONÍA PERUANA EN EL AÑO 2014”

TESIS
PARA OBTENER EL GRADO ACADÉMICO DE MAGISTER EN INGENIERÍA
DE SISTEMAS

Presentado por:

José Edgar García Díaz
Isaac Ocampo Yahuarcani

ASESOR: Ing. Carlos Alberto García Cortegano, Mg.

SAN JUAN BAUTISTA – PERÚ

2015

Eficiencia en las búsquedas de información sobre biodiversidad amazónica con el uso de un buscador semántico en los estudiantes de ciencias biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014

Presentado por:

JOSÉ EDGAR GARCÍA DÍAZ
ISAAC OCAMPO YAHUARCANI

Aprobado por:

Dr. Luis Benjamín Irigoín Sánchez

Presidente

MSc. Fernando Javier Salas Barrera

Miembro

Mg. Richard Alex López Albiño

Miembro

Mg. Carlos Alberto García Cortegano

Asesor

Dedicatoria

Dedicamos este trabajo a Dios y a todas las personas que creyeron en nosotros y nos dieron su apoyo incondicional y que a pesar de todas las adversidades de la vida nos demostraron que siempre estarán ahí a nuestro lado y que no nos dejaran solos, esto lo dedicamos a Ustedes.

Agradecimientos

Damos gracias a Dios, a la UNAP, a la Facultad de Ingeniería de Sistemas e Informática, a nuestras familias, a nuestro asesor y a nuestros hijos por comprendernos, ya que tuvimos que tomarnos sus tiempos para poder dedicarnos a desarrollar este trabajo y queremos que sepan que es para ellos a quienes va dirigido este trabajo.

Tabla de contenido

Dedicatoria.....	iii
Agradecimientos	iv
Tabla de contenido	v
ÍNDICE DE TABLAS	1
ÍNDICE DE FIGURAS	2
ÍNDICE DE GRÁFICOS	3
Resumen	4
Abstract	5
I. Introducción.....	6
1.1. Justificación de la Investigación.....	6
1.2. Identificación y formulación del problema	6
1.2.1. Problema general	7
1.2.2. Problema específico	8
1.3. Objetivos	8
1.2.3. Objetivo general.....	8
1.2.4. Objetivos específicos.....	8
II. Antecedentes	9
2.1. Investigaciones relacionadas al estudio.....	9
2.2. Marco teórico.....	13
2.2.1. Base de conocimiento.....	13
2.2.2. Buscador semántico	13
2.2.3. Buscador web	13
2.2.4. Motor de inferencia	14
2.2.5. Ontología	15
2.2.6. Web semántica	20
2.2.7. Eficiencia en las búsquedas	21
2.2.8. Información sobre biodiversidad.....	21
2.2.9. Usabilidad.....	21
2.3. Marco conceptual	23
2.4. Variables	25
2.4.1. Variable independiente (X)	25

2.4.2. Variable dependiente (Y).....	25
2.5. Indicadores e índices	26
2.6. Hipótesis.....	28
2.6.1. Hipótesis alterna.....	28
2.6.2. Hipótesis nula	28
III. Metodología	29
3.1. Tipo de investigación.....	29
3.2. Diseño de investigación.....	29
3.3. Población y muestra.....	29
3.3.1. Población	29
3.3.2. Muestra	29
3.4. Procedimiento, técnicas e instrumentos de recolección de datos.....	30
3.5. Procesamiento de la Información	31
3.6. Metodologías de desarrollo del software	31
3.6.1. Metodología XP	31
3.6.2. Método para creación de ontología.....	36
3.7. Herramientas para desarrollo de software.....	41
3.7.1. Lenguaje de programación PHP.....	41
3.7.2. MySql	42
3.7.3. Lenguaje RDF	43
3.7.4. Protégé.....	44
3.8. Desarrollo de la investigación	47
3.8.1. La estructura del pre test y post test	47
3.8.2. Implementación del cuestionario a los participantes	47
IV. Resultados	49
4.1. Evaluación de la implementación y uso del buscador semántico.....	49
4.2. Evaluación de la eficiencia en los resultados de búsqueda de información sobre biodiversidad amazónica	53
4.3. Tamaño del efecto en los resultados de las búsquedas de información sobre biodiversidad amazónica al aplicar el buscador simple y semántico.....	56
4.4. Contrastación de la hipótesis.....	56
4.4.1. Prueba t student	58
V. Discusión.....	61

VI. Conclusiones y recomendaciones	64
6.1. Conclusiones	64
6.2. Recomendaciones.....	65
Referencias bibliográficas.....	66
Anexos.....	71
ANEXO N° 01: Matriz de consistencia.....	72
ANEXO N° 02: Cuestionarios	73

ÍNDICE DE TABLAS

Tabla N°1: Evaluación del uso del buscador semántico realizado por los estudiantes de la facultad de ciencias biológicas de la UNAP – 2014.....	49
Tabla N°2: Evaluación pre test de la eficiencia de búsqueda de información sobre biodiversidad realizada al buscador simple, por los estudiantes de la facultad de ciencias biológicas de la UNAP – 2014.....	53
Tabla N°3: Evaluación post test de la eficiencia de búsqueda de información sobre biodiversidad realizada al buscador semántico, por los estudiantes de la facultad de ciencias biológicas de la UNAP – 2014.....	54
Tabla N°4: Resultados del pre test o buscador simple.....	56
Tabla N°5: Resultados del post test o buscador semántico.....	57
Tabla N°6: Presentación del calificativo.....	57
Tabla N°7: Definición de variables.....	58
Tabla N°8: Contrastación entre pre test y post test.....	58
Tabla N°9: Resultados obtenidos.....	59

ÍNDICE DE FIGURAS

Figura N° 1: Ciclo de vida de metodología XP.....	32
Figura N° 2: Interfaz de inicio del buscador.....	34
Figura N° 3: Interfaz de búsqueda con buscador simple.....	35
Figura N° 4: Interfaz de búsqueda con buscador semántico	35
Figura N° 5: Ontología básica sobre biodiversidad.....	40
Figura N° 6: Creación de clases de la ontología básica sobre biodiversidad utilizando Prótegé.....	45
Figura N° 7: Creación de instancias de la ontología básica sobre biodiversidad utilizando Prótegé.....	45
Figura N° 8: Distribución t student.....	59

ÍNDICE DE GRÁFICOS

Gráfico N° 1: Puntaje promedio de la valoración de indicadores del uso del buscador semántico.....	50
Gráfico N° 2: Puntaje promedio de los indicadores de evaluación del uso del buscador semántico.....	51
Gráfico N° 3: Evaluación promedio de la valoración de los indicadores del uso del buscador semántico.....	51
Gráfico N° 4: Evaluación promedio de los indicadores del uso del buscador semántico.....	52
Gráfico N° 5: Comparación entre los promedios de los indicadores de eficiencia del buscador simple y buscador semántico.....	55
Gráfico N° 6: Comparación entre los promedios de la valoración de los indicadores de eficiencia del buscador simple y buscador semántico	55

Resumen

La presente investigación tuvo como objetivo implementar un buscador semántico para incrementar la eficiencia en los resultados de búsqueda sobre términos de biodiversidad amazónica y fue realizado durante el periodo de enero a diciembre del año 2014. La investigación pertenece al tipo cuantitativo descriptivo y es de diseño cuasiexperimental, la población estuvo conformada por 250 estudiantes de la facultad de ciencias biológicas de la Universidad Nacional de la Amazonía Peruana y la muestra estuvo representada por 50 estudiantes elegidos por conveniencia, la técnica que se empleó en la recolección de datos fue la encuesta y el instrumento el cuestionario.

El procesamiento de datos se efectuó en forma manual y computarizada utilizando el programa estadístico SPSS versión 16.0 en español. El análisis e interpretación de la información se efectuó utilizando la estadística descriptiva, promedio, porcentaje y la estadística inferencial t – student.

Para la contrastación de la hipótesis principal se aplicó la prueba estadística inferencial paramétrica t – student, obteniendo t calculado $T_C = -7.524$, valor crítico $T_\alpha = -1.895$ con grado de libertad $gl = 7$, nivel de error de 5% $\alpha = 0.05$, es decir $T_\alpha > T_C$ rechazando así la hipótesis nula (H_0) y la hipótesis alterna es aceptada (H_α) siendo la conclusión que con la implementación de un buscador semántico se incrementó significativamente la eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014.

Palabras claves: semántica, eficiencia de resultados, biodiversidad

Abstract

The present research aimed to implement a semantic search engine for increased efficiency in search results on terms of Amazonian biodiversity and was conducted during the period from January to December 2014. The research belongs to the quantitative descriptive and quasi-experimental design the population consisted of 250 students of the Faculty of Biological Sciences, National University of the Peruvian Amazon and the sample was represented by 50 students chosen for convenience, the technique used in data collection was the survey and the instrument the questionnaire.

Data processing is done manually and computerized using SPSS version 16.0 in Spanish. The analysis and interpretation of the data was performed using descriptive statistics, average rate and inferential statistics t – student.

For the testing of the main hypothesis inferential statistical test was applied parametric t - student, obtaining t calculated $T_c = -7.524$, critical value $T_\alpha = -1.895$ with degrees of freedom $gl = 7$, error level of 5% $\alpha = 0.05$, then $T_\alpha > T_c$ rejecting the null hypothesis (H_0) and the alternative hypothesis is accepted (H_a) and the conclusion was that with the implementation of a semantic search engine is significantly increased efficiency in search results on Amazonian biodiversity in students of Biological Sciences, National University of the Peruvian Amazonia in 2014.

Keys words: semantic, efficiency in results, biodiversity

I. Introducción

1.1. Justificación de la Investigación

La sociedad peruana viene incorporando el uso de las TIC como mecanismos de producción, gestión y publicación de información en internet. Ello a la vez ha generado un gran desbordamiento en la generación de la cantidad de archivos relacionados a diversas temáticas, generándose a la vez que existan grandes volúmenes de información no relevante o inexacta, junto a la información útil para los usuarios. Siendo esta la causa de que este posea actualmente problemas de eficiencia en las búsquedas.

Por lo que es necesaria la existencia de herramientas que permitan incrementar la eficiencia de la búsqueda de información. Revisando las opciones tecnológicas actuales, encontramos en los mecanismos vinculados a los conceptos de la Web Semántica como un conjunto de mecanismos que podrían proporcionar medios para reducir esta problemática. Por lo que esta tesis plantea desarrollar un mecanismo tecnológico y a la vez realizar estudios comparativos entre la forma tradicional de búsqueda y un buscador semántico. Teniendo como temática relevante la biodiversidad amazónica peruana, siendo este un campo relevante en el desarrollo de la Amazonía.

Con todo ello se espera que el buscador semántico permita realizar búsquedas de mayor aproximación, interpretando la esencia de lo que busque el usuario. En esta etapa se ha priorizado lo que es la búsqueda de palabras, sin embargo en el mediano plazo se espera respuestas a partir de del significado de las frases, lo que incrementaría notablemente la potencia de la herramienta.

1.2. Identificación y formulación del problema

Desde los inicios de la web, uno de los principales servicios de mayor demanda y uso de internet lo constituye la búsqueda de información temática o académica (en este caso búsqueda de información sobre biodiversidad). Así mismo según el portal mexicano Publiometro (ssgt.com.mx/) “cada día las personas están expuestas a información equivalente a 174 periódicos de 85 páginas cada uno, cuando en 1980 eran 40 diarios producida por aproximadamente un millón de periódicos, 100 000 agencias de noticias, más de 1000 canales de televisión y millones de páginas de Internet en el mundo. Actualmente hay tanta producción de información que las búsqueda se hace complicado que un usuario encuentre la información que está buscando”. Para el caso de información sobre

biodiversidad amazónica del Perú, existen dos bancos relevantes el de la Plataforma Nacional de la Biodiversidad del Ministerio del Ambiente que cuenta con 600 mil registros de especies, el segundo caso lo constituye el Sistema de Información de la Diversidad Biológica y Ambiental de la Amazonía Peruana (www.siamazonia.org.pe) que cuenta con 44 mil registros de flora y fauna complementados por otros 40 mil registros de audio y video.

Es importante también mencionar que la web mundial cada 8 meses duplica su tamaño (www.alexa.com año 2011) y ello a la vez viene generando ciertos problemas que la vuelven ineficiente como son la “infoxicación”, entendida como la sobrecarga de información y las búsquedas limitadas o ineficientes alrededor de la denominada web superficial (viendo a la web superficial como el universo que puede ser indexado por los buscadores que según Piscitelli¹ 500 veces inferior a la web profunda). Los resultados que devuelven los buscadores no son precisamente los que buscan usuarios y a la vez demandan consumo de tiempo.

Así mismo, un problema relevante para acceder a información lo constituye una serie de limitaciones de buscadores como Google y Yahoo que sólo pueden recorrer la denominada web superficial que es una parte minúscula de la web (ello por las características informáticas de diversas plataformas web). Ello limita o proporciona escasa visualización de los registros de especies de Siamazonía en Google o Yahoo. A partir de ello se hace necesario contar con herramientas informáticas que faciliten mejoren la eficiencia de las búsquedas de información (en las plataformas actuales y desde los buscadores), haciendo que el usuario reduzca su exposición a la infoxicación reduciendo sus tiempos y procesos de búsqueda de información y que a la vez le permita aprovechar de alguna forma la denominada web profunda.

1.2.1. Problema general

¿Cómo el uso de un buscador semántico incrementa la eficiencia en los resultados de las búsquedas de información sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014?

¹ Filósofo argentino, especializado en los nuevos medios.

1.2.2. Problema específico

- ¿Cómo es la evaluación del uso de un buscador semántico sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014.
- ¿Cómo es la eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014?
- ¿Cuál es el tamaño del efecto de la eficiencia en los resultados de las búsquedas de información sobre biodiversidad amazónica al utilizar el buscador semántico en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014?

1.3. Objetivos

1.2.3. Objetivo general

Usar un buscador semántico para incrementar la eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014.

1.2.4. Objetivos específicos

- Evaluar el uso de un buscador semántico sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014.
- Evaluar la eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014.
- Determinar el tamaño del efecto de la eficiencia en los resultados de las búsquedas de información sobre biodiversidad amazónica al utilizar el buscador semántico en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014.

II. Antecedentes

2.1. Investigaciones relacionadas al estudio

- Criado (2009), ha desarrollado su tesis doctoral denominada “Procedimiento semi – automático para transformar la Web en Web Semántica” que tuvo como objetivo desarrollar un procedimiento para contribuir en la extensión o crecimiento de las ontologías basado en la conversión de contenidos de la Web a contenido Semántico. En consecuencia, al transformar un sitio web normal en un sitio web semántico se obtendrán funcionalidades semánticas que podrán ser explotadas por un buscador semántico y cuando este sea tratado por un buscador normal todo ello será normal. El proceso comprendió un proceso de identificación o análisis de la ontología del sitio web, con la herramienta se realizó el proceso de forma automática, posteriormente se trabajó en el procesamiento de lenguaje natural, la generación de vistas semánticas, se trabajó con un grupo de usuarios encargados de verificar las funcionalidades y los resultados. Se ha validado el funcionamiento de la herramienta sin embargo si en el futuro se quisiera lograr su masificación será necesaria avanzar en mejorar las técnicas de procesamiento de lenguaje natural.
- Lozano (2002), ha desarrollado su tesis doctoral denominada “Métrica de idoneidad de ontologías que tuvo como objetivo definir y validar una metodología que permite medir la idoneidad de las ontologías con respecto al contexto de sus sistemas (Web Semántica). Este interés surge porque a la fecha de elaboración (2002), no existe ningún método que permita contextualizar de mejor manera las ontologías. Esta metodología ha sido denominada Ontometric. Una vez organizada la información, usuarios experimentados en ontologías valoran los niveles o grados de importancia de cada característica a través de la aplicación de un cuestionario de preguntas (además se indican los defectos y omisiones). Entre los principales resultados se tiene que se ha validado un método que valora cuantitativamente la idoneidad de las ontologías, lo que permite una serie de condiciones favorables para otros procesos de eficiencia en el sistema.
- Muñoz (2009), ha desarrollado su tesis doctoral denominada “Teoría de modelado del E-learning y aplicación a un sistema de pistas adaptativo en tutoría inteligente utilizando técnicas de Web Semántica” que tuvo como propósito establecer y validar una solución de tutoría inteligente basado en conceptos relacionados a e-learning y Web Semántica, mediante el uso de pistas en el aprendizaje basado en problemas a través del

computador. Este trabajo proporciona una teoría de modelado de e-learning que incluye una visión global sobre qué modelar y cómo hacerlo, las interrelaciones entre diferentes conceptos y elementos, así mismo. Además se ha definido una nueva arquitectura que permite combinar sistemas de tutoría inteligente existentes con técnicas de Web Semántica, explicando sus elementos, relaciones, retos y diferentes criterios de diseño. Esta arquitectura definida permite incorporar las ventajas de la web semántica dentro de los sistemas de tutoría inteligente. Para poner en práctica el modelo de pistas se ha desarrollado una aplicación de pistas. Para la evaluación del modelo de pistas, se han realizado diferentes experiencias de un total de tres asignaturas, en una de ellas durante cuatro ediciones. Se han tomado datos cuantitativos y cualitativos globales de la herramienta y las interacciones de los alumnos, se ha comparado el incremento de aprendizaje con respecto a un sistema de evaluación sin pistas, y con respecto a tutores humanos generando pistas. Los resultados sobre estos aspectos concluyen un efecto positivo del modelo de pistas y del sistema reproductor de pistas implementados. Además se ha analizado qué estrategias de provisión de pistas producen mayores incrementos del aprendizaje y cómo diferentes técnicas de provisión de pistas producen diferentes comportamientos de los alumnos.

- Navarro (2012), ha desarrollado su tesis doctoral denominada “FLERSA: Un sistema semántico de gestión de contenido Web (S-CMS)” que tuvo como objetivo definir y desarrollar el concepto de S-CMS para reducir en parte las carencias de contenido semántico de la web actual y aportar a los CMS los beneficios de la Web Semántica, adaptando la Web Semántica a los sistemas actuales. Se desarrolla un prototipo que comprende en la definición requisitos iniciales de diseño, la arquitectura semántica, rol de la ontología, luego se realiza pruebas del proceso manual y automático. A partir del cual se realiza un estudio comparativo de las funcionalidades que aporta la herramienta con respecto a otras herramientas (se hacen la prueba con otras 4 herramientas de la misma funcionalidad). Entre los resultados más relevante podemos comentar que se ha demostrado que el sistema permite incorporar Web Semántica a los sistemas informáticos convencionales, sin embargo cabe indicar que aún es necesario realizar ciertos niveles de esfuerzos en los sistemas informáticos.
- Noriega (2009), ha desarrollado su tesis doctoral denominada “Modelado de análisis de sistemas CSCW siguiendo el enfoque de Ingeniería dirigida por Ontologías” que tuvo como objetivo establecer los fundamentos de la adopción de un análisis y modelado de los sistemas Computer Supported Cooperative Work-CSCW o grupos de trabajo

colaborativo siguiendo el enfoque de ingeniería dirigida, además del empleo de ontologías, comprendiendo una serie de actividades (recreativo, ocio, juego). Se ha realizado proceso de adaptación, se ha elaborado e implementado la ontología y se han realizado procesos validación que ha permitido determinar la validez y utilidad del modelado, teniendo en cuenta la necesidad de una serie de condiciones.

- Palacios (2010), ha desarrollado su tesis doctoral denominada “Sistema de recuperación conceptual mediante niveles semánticos en la representación de esquemas de metadatos” que tuvo como objetivo desarrollar un plano semántico a esquemas de metadatos y sus documentos derivados, mediante la relación de sus elementos con conceptos definidos en un recurso semántico compartido vía Web. Una vez definida la ontología y las condiciones relacionadas al software, se analizarán los métodos existentes y se seleccionará aquel que mejor se adecue a la propuesta, esto en base a criterios de extensibilidad, modificabilidad, navegabilidad, accesibilidad. Una vez seleccionado el método de correspondencia, se definirá el modo de articularlo, tomando como referencia los criterios de extensibilidad, compatibilidad y reutilización. El desarrollo de la propuesta ha logrado verificar y validar la metodología para la cualificación semántica de esquemas y su correspondencia con la ontología de referencia. Además se incluye la especificación del sistema informático que soportará los procesos de interconexión y gestión tanto de esquemas como de ontologías.
- Reyes (2007), ha desarrollado su tesis doctoral denominada “Propuesta para el modelado del conocimiento empresarial” que tuvo como objetivo desarrollar un marco de referencia que ayude a las pequeñas y medianas empresas virtuales a realizar modelos empresariales que puedan integrarse y ser útiles en su sistema de gestión del conocimiento, basados en la reducción de los problemas de interoperabilidad relacionados con el modelado del conocimiento empresarial. Para ello se ha empleado la metodología KM-IRIS para lo cual se ha identificado el conocimiento objetivo y las variables de entrada y fuentes necesarias para su extracción, se ha analizado los siguientes bloques Organización, proceso, producto/servicio y recurso, producto de ello se han obtenido las diferentes categorías y subcategorías ontológicas. Entre las conclusiones más relevantes de la tesis se tiene que el uso de las TIC es necesario pero no suficiente para desarrollar las empresas, siendo necesariamente primordial el intercambio de información entre las empresas y los partners.
- Romero (2007), ha desarrollado su tesis doctoral denominada “Especificación OWL de una ontología para teleeducación en la Web Semántica” que tuvo como objetivo la

especificación y validación de una ontología en el marco de la Web Semántica como insumo relevante para teleeducación. El proceso comprendió en la generación de una ontología base, luego se realiza la prueba de validación, así mismo se desarrollan una serie de prototipos que permitirá realizar una prueba de concepto sobre la ontología definida, luego se compruebo la corrección lógica de la misma por medio de la clasificación y el chequeo de consistencia. Luego se realiza una prueba del uso por medio de agente inteligente, evaluando primero los principales razonadores o motores de inferencia actuales y modelizando después la carga de conocimientos. También se realizan pruebas de búsquedas por palabras clave o búsquedas por medio de inferencias. Todo esto permitirá validar que la ontología es implementable y utilizable por medio de agentes inteligentes.

- Torres (2008), ha desarrollado su tesis doctoral denominada “Sistemas de análisis automático de fotografías. Modelo conceptual según los estándares de la Web Semántica” que tuvo como objetivo comprobar que es posible la recuperación de fotografías a través de la Web Semántica, además de conocer si es que existen estándares suficientemente completos para que este proceso pueda ser realizado en el futuro sin dificultades. Teniéndose como resultado que el análisis de las fotografías automáticamente sí es posible, siendo necesario para lograr ello una serie de procesos de indexación previos que permitan que el sistema aprenda a correlacionar los contenidos visuales con los conceptos. Además se ha comprobado que actualmente existen estándares que harán posible la integración del análisis automático de las fotografías en la Web Semántica en el futuro.

2.2. Marco teórico

2.2.1. Base de conocimiento

Según Romero (2007), es el conjunto de sentencias, que pueden ir creciendo y variando en el tiempo, en un lenguaje formal, de forma que los agentes inteligentes reunirán las sentencias de otros servicios en la web, y construirán su base de conocimiento sobre la que trabajar. En la mayoría de los casos los agentes inteligentes son programados con programación declarativa, en la que el agente obtiene la información que necesita saber y se le plantea que resuelva un problema, sin indicarle el algoritmo determinado para resolver ese problema.

2.2.2. Buscador semántico

Un buscador semántico es aquel que realiza el rastreo atendiendo al significado del grupo de palabras que escribes y no basándose en las actuales etiquetas. En pocas palabras, un buscador inteligente. Aunque todavía su uso es incipiente, ya existen algunos que ofrecen resultados reseñables e interesantes y ya son referencia para el futuro de la búsqueda de información (Corneliu, 2005).

Los buscadores semánticos son un ejemplo más de aplicaciones basadas en Web Semántica. El objetivo es satisfacer las expectativas de búsqueda de usuarios que requieren respuestas precisas (Sitio Web Buscadores Semánticos de Guiomar Pérez López, 2010).

Un buscador semántico efectúa la búsqueda atendiendo al significado del grupo de palabras que escribes sin basarse en las actuales etiquetas. Sintetizando, podríamos decir que se trata de un Buscador Inteligente (Sitio Web Acción TI, 2003)

2.2.3. Buscador web

Los buscadores son una herramienta imprescindible para recuperar la ingente cantidad de información diseminada por Internet. Todas las páginas conocidas genéricamente como buscadores devuelven un listado de páginas web cuando se introduce un término de búsqueda, pero no todas lo hacen de la misma forma (Portal de la Fundación Eroski).

Un motor de búsqueda, también conocido como buscador, es un sistema informático que busca archivos almacenados en servidores web. Un ejemplo son los buscadores de Internet (algunos buscan únicamente en la web, pero otros lo hacen además en noticias, servicios como Gopher, FTP, etc.) cuando se pide información sobre algún tema. Las búsquedas se hacen con palabras clave o con árboles jerárquicos por temas; el resultado de la búsqueda es un listado de direcciones web en los que se mencionan temas relacionados con las palabras clave buscadas (sitio web informática 1101).

La gran mayoría de los buscadores se basan en palabras clave. Es decir, el usuario introduce las palabras en su búsqueda y la aplicación devuelve todos los documentos que contienen esas palabras. Estos buscadores presentan fundamentalmente las siguientes desventajas:

- Escasa precisión o relevancia en los resultados: esto significa que se devuelven muchos documentos poco relevantes para la búsqueda.
- Debemos tener en cuenta que la presencia de una palabra clave en un documento no implica que éste sea relevante para nosotros.
- Excesiva sensibilidad al vocabulario empleado en las búsquedas: y por lo tanto, la dificultad de obtener a la primera todos los resultados relevantes disponibles.
- Muchos documentos de interés pueden no incluir las palabras clave, pero sí sinónimos, hipónimos o hiperónimos de ellas.

2.2.4. Motor de inferencia

Según Romero (2007), es lo que permite sacar conclusiones de la base de conocimiento y resolver el problema que se le ha planteado. El motor de inferencia permitirá obtener nuevas sentencias (conclusiones) a partir de las sentencias de su base de conocimiento, y determinar, por ejemplo, si una sentencia es deducible a partir de las sentencias de su base de conocimiento.

2.2.5. Ontología

Para Castells (2003), “una ontología es una jerarquía de conceptos con atributos y relaciones, que define una terminología consensuada para definir redes semánticas de unidades de información interrelacionadas”. En el caso de los autores López y García (2005), “una ontología no es más que una especificación de lo que existe en un dominio, convirtiéndose éstas en una pieza fundamental de las tecnologías orientadas a la web”. Por su parte Uschold y Grüninger (1996), definen ontología como “una comprensión compartida de algún dominio de interés. Aunque una definición más amplia e informal es la que define ontología como un vocabulario de términos y alguna especificación de su significado”. Gruber (2003), también añade que la “ontología es una especificación explícita de una conceptualización”. También López y García (2005), comenta algo muy importante contextualizando las ontologías desde el punto de vista de la ingeniería lo define de la siguiente manera: “desde un punto de vista de ingeniería una ontología es un artefacto que está constituido por un vocabulario específico utilizado para describir una cierta realidad, además de un conjunto de suposiciones que tienen en cuenta el significado del vocabulario”. Y entre las facilidades que otorgan a los desarrolladores su uso está en que permite compartir un entendimiento común de una estructura de información entre personas o agentes software, posibilitar la reutilización de conocimiento de un dominio, hacer explícitas suposiciones de un dominio, separar el conocimiento de un dominio del conocimiento operacional, o analizar el conocimiento de un dominio (Noy & McGuinness, 2001).

La creación de una ontología ofrece la ventaja de que se hace explícita la categorización de los elementos y relaciones que intervienen en el modelo de conocimiento, de forma que, por un lado, el modelo de conocimiento puede ser editado y gestionado y, por otro, es posible transmitirlo de manera que un sistema entienda la conceptualización que se ha utilizado en otro. Las ontologías proporcionan un conocimiento del dominio reusable y mantenible (Mizoguchi et al., 1997).

Muchos autores coinciden a la vez que una de las principales limitaciones a la hora de realizar la web semántica en la práctica es la de consensuar ontologías en una comunidad por poco amplia que sea. Converger a una representación común es una labor más compleja de lo que puede parecer, ya que típicamente cada parte del sistema con lleva peculiaridades necesarias, y un punto de vista propio que a menudo necesitan incidir en la propia ontología (Castells, 2003).

“Una ontología proporciona un vocabulario de clases y relaciones para describir un dominio, además de ser una especificación formal y explícita de una conceptualización compartida”, esto se entiende mejor de la siguiente manera (Studer, 1998):

- Formal: Se refiere al hecho de que la ontología debe ser legible por un ordenador, excluyendo el lenguaje natural.
- Explícita: Significa que los conceptos que utilizan sus limitaciones se definen explícitamente.
- Conceptualización: Se refiere a la identificación de los conceptos más relevantes de un fenómeno del mundo.
- Compartida: Quiere decir que una ontología captura un consenso de conocimiento, es decir que el conocimiento no proviene de un solo individuo sino que es aceptada por un grupo.

Respecto a los tipos de ontologías también existen definiciones y aseveraciones de una infinidad de autores sin embargo en el paper titulado “Web Semántica y Ontologías” de Francisco José García Peñalvo (2007), presenta una recopilación de clasificaciones como la de Uschold (1996), que menciona tres tipologías mayores a partir de factores como la formalidad, el propósito y la materia relacionada al origen de la ontología. En cuanto a la formalidad define los siguientes:

- Ontología altamente informal: Expresada en lenguaje natural teniendo como ejemplos a los glosarios de términos).
- Ontología informal estructurada: Expresada en lenguaje natural estructurado y restringido.
- Ontología semiformal: Usa un lenguaje de definición formal, como puede ser ontolingua (Gruber, 1993).
- Ontología rigurosamente formal: La definición de términos se lleva a cabo de manera meticulosa usando semántica formal, teoremas, y pruebas de estas propiedades como solidez y completación (Grüninger & Fox, 1995b).

En cuanto al propósito se definen los siguientes:

- Ontologías para comunicación entre personas: Una ontología informal no ambigua puede ser suficiente.
- Ontologías para interoperabilidad entre sistemas: Para llevar a cabo traducciones entre diferentes métodos, lenguajes, software. En estos casos la ontología se usa como un formato de intercambio de conocimiento.
- Ontologías para beneficiar la ingeniería de sistemas: Cuando las ontologías benefician las aplicaciones software apoyando aspectos como la reutilización de componentes software en un dominio de interés, la adquisición de conocimiento, la fiabilidad de los sistemas al proporcionar consistencia en el conocimiento utilizado, o la especificación de los sistemas software identificando los requisitos y definiendo especificaciones para las tecnologías de la información.

En cuanto a la materia se definen los siguientes:

- Ontologías de dominio: Caracterizan disciplinas específicas, tales como medicina, finanzas, química, biología con independencia de los problemas o tareas relevantes de dichas disciplinas.
- Ontologías de resolución de problemas: Conceptualizan el problema o la tarea a resolver en un dominio.
- Ontologías de representación o metaontologías: El objeto que se caracteriza es un lenguaje de representación de conocimiento.

Así mismo Contreras y Martínez en su tutorial de Ontologías (2007), destacan las siguientes metodologías para construir ontologías:

- Diligence: Basado en la colaboración de múltiples participantes, de manera que la creación de una ontología se concibe como un proceso social, distribuido y muy poco controlado. Se hace hincapié en el consenso, facilitando un marco para la discusión de diferentes propuestas y en el intercambio de argumentos a favor de las diferentes posibilidades hasta llegar a un acuerdo final.

- **Competency Questions:** Comprende la determinación del dominio y el alcance de la ontología mediante la lista de preguntas que el sistema debería ser capaz de contestar. Las respuestas a estas preguntas sugieren lo que podrían ser las instancias de la ontología, a partir de las cuales se deducirían (generalizando) las clases de la misma. Al mismo tiempo, estas preguntas servirán como factor decisivo a la hora de evaluar la propia ontología.
- **Methontology:** Propone un ciclo de vida de construcción de la ontología basado en prototipos evolutivos, porque esto permite agregar, cambiar y remover términos en cada nueva versión (prototipo). Para cada prototipo, el proceso consta de los siguientes pasos esenciales: especificación, conceptualización, formalización, implementación y mantenimiento.
- **On-To-Knowledge:** Se centra en las aplicaciones futuras de la ontología a la hora de diseñarla. Los pasos esenciales de que consta son: estudio de viabilidad, delimitación del dominio y objetivo, evaluación y mantenimiento.

En donde creemos ha existido mejor consenso por parte de los autores es en la determinación de los componentes de las ontologías, donde se ha identificado lo siguiente (Sowa, 2000; Noy & McGuinness, 2001; Farquhar, 1997):

- **Axioma:** Elementos que permiten el modelado de verdades que se cumplen siempre en la realidad. Los axiomas pueden ser estructurales y/o no estructurales.
- **Clase o tipo:** Una clase es un conjunto de objetos (físicos, tareas, funciones...). Cada objeto en una clase es una instancia de esa clase. Desde el punto de vista de la lógica los objetos de una clase se pueden describir especificando las propiedades que éstos deben satisfacer para pertenecer a esa clase. Las clases son la base de la descripción del conocimiento en las ontologías ya que describen los conceptos del dominio. Una clase puede ser dividida en subclases, las cuales representarán conceptos más específicos que la clase a la que pertenecen. Una clase cuyos componentes son clases, se denomina superclase o metaclase.
- **Instancias o individuos:** Son objetos, miembros de una clase, que no pueden ser divididos sin perder su estructura y características funcionales. Pueden ser agrupados en clases.

- Relaciones: Se establecen entre conceptos de una ontología para representar las interacciones entre éstos. Definidas por lo general como el producto cartesiano de n conjuntos: $R: C_1 \times C_2 \times \dots \times C_n$. Algunas de las relaciones más utilizadas son:
 - Instancia de: Asocian objetos a clases.
 - Relaciones temporales: Implican precedencia en el tiempo.
 - Relaciones topológicas: Establecen conexiones espaciales entre conceptos.

- Propiedades o *slots*: Los objetos se describen por medio de un conjunto de características o atributos que son almacenados en los *slots*. Éstos almacenan diferentes clases de valores. Las especificaciones, rangos y restricciones sobre estos valores se denominan características o facetas.

- Frame: Un objeto que incluye clases, instancias y relaciones.

- Conceptualización: Conjunto de conceptos, relaciones, objetos y restricciones que caracterizan un dominio.

- Taxonomía: Conjunto de conceptos organizados jerárquicamente. Las taxonomías definen las relaciones entre los conceptos, pero no los atributos de éstos.

- Vocabulario: Conjunto de palabras con una explicación y documentación que persigue.

Por otro lado la tecnología que se ha creado para hacer posible la web semántica incluye lenguajes para la representación de ontologías, parsers, lenguajes de consulta, entornos de desarrollo, módulos de gestión (almacenamiento, acceso, actualización) de ontologías, módulos de visualización, conversión de ontologías, y otras herramientas y librerías (Castells, 2003). Entre los que destacan:

- Lenguajes de Definición o Representación: Se utilizan para definir o estructurar las ontologías. Destacan entre los principales el RDF (Resource Description Framework) que es un lenguaje que permite describir formalmente el significado de la terminología usada en los documentos Web, y el OWL (Web Ontology

Language), es un lenguaje de marcado para publicar y compartir datos usando ontologías en la web.

- Editores de Ontologías: Herramientas informáticas que se usan para la elaboración de Ontologías, están basados en la infinidad de lenguajes de definición.
- Lenguajes de Consulta: Se utilizan para realizar búsquedas lógicas sobre las ontologías.
- Motor de Inferencia: Aplicación informática que permite realizar consultas lógicas, sirve para que las computadoras realicen el proceso de razonamiento humano.

Parser: Aplicación informática que permite realizar análisis sintáctico de las cadenas de texto en las que está escrita la Ontología. Verifica que las etiquetas estén bien escritas.

2.2.6. Web semántica

El consorcio World Wide Web Consortium - W3C, que dicta las recomendaciones técnicas para la web, que incluso es liderado por el creador de la Web Tim Bernes - Lee, define a la web semántica como “una web extendida, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida. Al dotar a la web de más significado y, por lo tanto, de más semántica, se pueden obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, es posible compartir, procesar y transferir información de forma sencilla. Esta web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante” (W3C, 2008). Además es importante disgregar que la web semántica no es una nueva web segregada de la actual. Es una extensión de la actual en la que la información se ofrece con un significado bien definido, permitiendo a ordenadores y personas a trabajar de forma cooperativa (Berners-Lee et al., 2001; Hendler et al., 2002). Quizás la idea central de la web semántica respecto al usuario final sea “disponer datos en la web definidos y enlazados de forma que puedan ser utilizados por las maquinas no solamente para

visualizarlos sino también para: automatizar tareas, integrar y reutilizar datos entre aplicaciones (Berners-Lee, 2005)".

Romero (2007), en su tesis doctoral define la web semántica como una meta-web que se construye encima de la web actual. Se fundamenta en la idea de tener datos en la web definidos y vinculados de forma que puedan ser usados por máquinas no sólo a la hora de pintarlo sobre pantalla, sino para tareas de automatización, integración, interpretación y reutilización de datos a través de varias aplicaciones. El objetivo de la web semántica es proveer un marco común que permita que los datos sean compartidos y reutilizados por distintas aplicaciones, compañías y comunidades. Los sistemas de web semánticas actuales se basan en una base de conocimiento y algún motor de inferencia.

2.2.7. Eficiencia en las búsquedas

Según Beltré (2008), es un sistema “la eficiencia en uso es el nivel de productividad del usuario cuando utiliza el sistema; se puede medir en base al número de tareas por unidad de tiempo que el usuario (experto) es capaz de llevar a cabo utilizando el sistema. La tasa de errores se refiere a los errores cometidos durante el uso del sistema y cuán fácilmente el usuario se puede recuperar de ellos, tanto del número como del tipo de errores; se puede medir en base al número de errores que el usuario comete cuando intenta llevar a cabo una tarea concreta, y cómo se recupera del error”. Así mismo indica que la eficiencia comprende los tiempos requeridos por los empleados para desarrollar las tareas se reducen aumentando su productividad.

2.2.8. Información sobre biodiversidad

Según Siamazonía² (IIAP³), se considera información sobre biodiversidad al “conjunto de registros de datos relacionados a flora y fauna de la Amazonía peruana (incluye audio, video, presentaciones), organizado de acuerdo a algún esquema”.

2.2.9. Usabilidad

Según la Organización Internacional para la Estandarización la usabilidad es “la

² Sistema de Información de la Diversidad Biológica y Ambiental de la Amazonia Peruana

³ Instituto de Investigaciones de la Amazonía Peruana - IIAP

eficacia, eficiencia y satisfacción con la que un producto permite alcanzar objetivos específicos a usuarios específicos en un contexto de uso específico”.

Según Yusef Hassan & Francisco J. Martín Fernández & Ghzala Iazza, el concepto de usabilidad puede ser definido, además de como atributo de calidad de una aplicación, consecuentemente, como disciplina o enfoque de diseño y evaluación. Se suele hablar entonces de Ingeniería de la Usabilidad - conjunto de fundamentos teóricos y metodológicos que aseguren el cumplimiento de los niveles de usabilidad requeridos para la aplicación.

Un concepto íntimamente ligado al de usabilidad es el de accesibilidad. Éste ya no se refiere a la facilidad de uso, sino a la posibilidad de acceso. En concreto a que el diseño, como prerrequisito imprescindible para ser usable, posibilite el acceso a todos sus potenciales usuarios, sin excluir a aquellos con limitaciones individuales - discapacidades, dominio del idioma o limitaciones derivadas del contexto de acceso - software y hardware empleado para acceder, ancho de banda de la conexión empleada, etc.

En la definición podemos observar que la usabilidad se compone de dos tipos de atributos:

- **Atributos cuantificables de forma objetiva:** como son la eficacia o número de errores cometidos por el usuario durante la realización de una tarea, y eficiencia o tiempo empleado por el usuario para la consecución de una tarea.
- **Atributos cuantificables de forma subjetiva:** como es la satisfacción de uso, medible a través de la interrogación al usuario, y que tiene una estrecha relación con el concepto de usabilidad percibida.

Según ISO 9241-11⁴, las dimensiones de la usabilidad se fundamentan en:

- **Eficacia (effectiveness):** la exactitud y la cantidad con el cual los usuarios alcanzan metas especificadas.
- **Eficiencia (efficiency):** los recursos gastados con relación a la certeza con cual usuarios logran las metas (en este caso las búsquedas de información).

⁴ ISO (the International Organization for Standardization)

2.3. Marco conceptual

- **Buscador**

Plataforma o aplicación informática que presenta funcionalidades de indexación y búsquedas de contenidos en sitios web.

- **Búsqueda**

Según el documento “Programación I” de la Universidad Nacional del Nordeste de Argentina “la operación de búsqueda nos permite encontrar datos que están previamente almacenados. La operación puede ser un éxito, si se localiza el elemento buscado o un fracaso en otros casos. La búsqueda se puede realizar sobre un conjunto de datos ordenados, lo cual hace la tarea más fácil y consume menos tiempo; o se puede realizar sobre elementos desordenados, tarea más laboriosa y de mayor insumo de tiempo”.

- **Interoperabilidad**

La habilidad de dos o más sistemas o componentes para intercambiar información y utilizar la información intercambiada.

- **RDF (Resource Description Framework)**

Modelo de datos para objetos (recursos) y las relaciones entre ellos. Ya tiene la capacidad de expresar cierta semántica.

- **RDF Schema**

Vocabulario para la descripción de propiedades y clases de recursos RDF. Cuenta con semántica para la generalización de jerarquías de las propiedades de las clases.

- **Semántica**

Al hablar de la semántica y de las otras ciencias del epígrafe relacionadas con ella, nos estamos refiriendo al significado de las palabras, no al significado conjunto de un enunciado o de un texto en el que intervienen muchos factores, a veces difíciles de delimitar. Dentro de la semántica se encuentra la lexicología, que estudia las palabras pertenecientes al lenguaje general usado por todos, y en paralelismo, hallamos la

terminología, que estudia las palabras del lenguaje especializado, es decir, los términos usados para transmitir el conocimiento de un campo del saber.

- **Web**

World Wide Web (o la "Web") es un sistema de documentos de hipertexto y/o hipermedios enlazados y accesibles a través de Internet. Con un navegador Web, un usuario visualiza páginas web que pueden contener texto, imágenes, vídeos u otros contenidos multimedia, y navega a través de ellas usando hiperenlaces.

- **Web 2.0**

Comprende aquellos sitios web que facilitan el compartir información, la interoperabilidad, el diseño centrado en el usuario y la colaboración en la World Wide Web.

- **Web 3.0**

También llamada web semántica basada en contenidos web organizados o consultable por significados.

- **Web profunda**

Conjunto de páginas web de los sitios web, sobre lo que los buscadores no pueden realizar procesos de indexación de documentos (contenido no está relacionado con buscadores).

- **Web Superficial**

Conjunto de páginas web de los sitios web, sobre lo que los buscadores pueden realizar procesos de indexación de documentos (contenidos están relacionados con buscadores).

2.4. Variables

2.4.1. Variable independiente (X)

Buscador semántico.

2.4.2. Variable dependiente (Y)

Eficiencia de las búsquedas de información sobre biodiversidad.

2.5. Indicadores e índices

Variables	Indicadores	Índices
VARIABLE INDEPENDIENTE (X) Buscador Web Semántico	Sobre el tiempo de apertura de interfaz: ¿Considera usted que el tiempo para visualizar la interfaz del buscador es apropiado para los usuarios?	- Muy apropiado (5) - Ligeramente apropiado (4) - Apropiado (3) - Ligeramente inapropiado (2) - Muy inapropiado (1)
	Sobre el proceso a seguir para realizar una búsqueda: ¿Considera usted que la secuencia seguida en el buscador para realizar una búsqueda es apropiado para los usuarios?	- Muy apropiado (5) - Ligeramente apropiado (4) - Apropiado (3) - Ligeramente inapropiado (2) - Muy inapropiado (1)
	Sobre el tiempo de respuesta a la búsqueda: ¿Considera usted que el tiempo de respuesta del buscador para encontrar y listar los resultados es apropiado para los usuarios?	- Muy apropiado (5) - Ligeramente apropiado (4) - Apropiado (3) - Ligeramente inapropiado (2) - Muy inapropiado (1)
	Sobre el listado de registros de búsqueda: ¿Considera usted que el listado de registros (visibilidad y entendimiento del resultado) del buscador es apropiado para los usuarios?	- Muy apropiado (5) - Ligeramente apropiado (4) - Apropiado (3) - Ligeramente inapropiado (2) - Muy inapropiado (1)

Variables	Indicadores	Índices
<p>VARIABLE DEPENDIENTE (Y)</p> <p>Eficiencia de las búsquedas de información sobre biodiversidad.</p>	<p>Sobre el tiempo que utilizó para realizar en una búsqueda:</p> <p>¿Considera usted que el tiempo total que necesitó para realizar su búsqueda es apropiado?</p>	<ul style="list-style-type: none"> - Muy apropiado (5) - Ligeramente apropiado (4) - Apropiado (3) - Ligeramente inapropiado (2) - Muy inapropiado (1)
	<p>Sobre la búsqueda realizada:</p> <p>¿Considera usted que la cantidad de pasos realizados en el buscador desde la apertura de interfaz hasta encontrar los resultados es apropiado?</p>	<ul style="list-style-type: none"> - Muy apropiado (5) - Ligeramente apropiado (4) - Apropiado (3) - Ligeramente inapropiado (2) - Muy inapropiado (1)
	<p>Sobre la cantidad de registros encontrados:</p> <p>¿Considera usted que la cantidad de registros encontrados por el buscador es apropiado?</p>	<ul style="list-style-type: none"> - Muy apropiado (5) - Ligeramente apropiado (4) - Apropiado (3) - Ligeramente inapropiado (2) - Muy inapropiado (1)
	<p>Sobre la utilidad de los registros encontrados:</p> <p>¿Considera usted que los contenidos (información) de los registros encontrados es apropiado?</p>	<ul style="list-style-type: none"> - Muy apropiado (5) - Ligeramente apropiado (4) - Apropiado (3) - Ligeramente inapropiado (2) - Muy inapropiado (1)

2.6. Hipótesis

2.6.1. Hipótesis alterna

El uso de un buscador semántico incrementará significativamente la eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014.

2.6.2. Hipótesis nula

El uso de un buscador semántico no incrementará significativamente la eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014.

III. Metodología

3.1. Tipo de investigación

La investigación es cuantitativa del tipo descriptivo. Es descriptivo porque explica como es y cómo se manifiestan las variables Buscador Web Semántico y Eficiencia de las búsquedas de información sobre biodiversidad.

3.2. Diseño de investigación

La investigación pertenece al diseño cuasiexperimental. Los diseños cuasiexperimentales manipulan por lo menos una variable independiente para observar su efecto y relación con una o más variables dependientes.

El esquema es:

G₁: O₁ X O₂

Dónde:

G₁: Grupo experimental

X: Tratamiento con el prototipo

O₁: Test antes del experimento

O₂: Test después del experimento

3.3. Población y muestra

3.3.1. Población

La población está conformada por 250 alumnos matriculados en la facultad de Ciencias biológicas de la Universidad Nacional de la Amazonia Peruana en el semestre 2014 – II.

3.3.2. Muestra

La muestra está conformada por 50 estudiantes del Tercer nivel, V y VI ciclo matriculados en la facultad de Ciencias Biológicas en el semestre 2014 – II. La selección de la muestra se efectuará en forma no probabilística por conveniencia.

3.4. Procedimiento, técnicas e instrumentos de recolección de datos

3.4.1. Procedimiento de recolección de datos

Para el procedimiento de recolección de datos se realizó los siguientes pasos:

- Diseño del plan de tesis
- Revisión bibliográfica
- Elaboración del plan de tesis
- Diseño y elaboración del instrumento de recolección de datos
- Presentación del plan de tesis
- Validación y confiabilidad de los instrumentos de recolección de datos
- Aplicación de los instrumentos de la recolección de datos
- Procesamiento de la información
- Análisis e interpretación de la información
- Elaboración de discusión, conclusiones y recomendaciones
- Redacción del informe final
- Presentación, revisión, reajuste y aprobación del informe de tesis
- Sustentación de la tesis.

3.4.2. Técnica de recolección de datos

La técnica de recolección de datos para la variable independiente, Buscador Web Semántico, fue la encuesta y se observó los hechos en forma indirecta y para la variable dependiente, eficiencia de las búsquedas de información sobre biodiversidad, también fue la encuesta.

3.4.3. Instrumentos de recolección de datos

El instrumento de recolección de datos para la variable independencia Buscador Web Semántico y para la variable dependiente Eficiencia de las búsquedas de información sobre biodiversidad será el cuestionario el que será sometido a la prueba de validez y confiabilidad antes de su aplicación.

3.5. Procesamiento de la Información

La información fue procesada en forma computarizada empleando la herramienta SPSS versión 16.0 en español sobre la base de datos, cuyos resultados se organizaron en cuadros y representados en gráficos.

El análisis e interpretación de la información se efectuara mediante:

Distribución t-student.

Análisis Univariado: a través de la estadística descriptiva

Análisis Bivariado: a través de la estadística inferencial

3.6. Metodologías de desarrollo del software

3.6.1. Metodología XP

La metodología XP define cuatro variables para cualquier proyecto de software: **costo**, **tiempo**, **calidad** y **alcance**. Además, se especifica que, de estas cuatro variables, sólo tres de ellas podrán ser fijadas arbitrariamente por actores externos al grupo de desarrolladores (clientes y jefes de proyecto). El valor de la variable restante podrá ser establecido por el equipo de desarrollo, en función de los valores de las otras tres. Este mecanismo indica que, por ejemplo, si el cliente establece el alcance y la calidad, y el jefe de proyecto el precio, el grupo de desarrollo tendrá libertad para determinar el tiempo que durará el proyecto. Este modelo es analizado por Kent Beck, donde propone las ventajas de un contrato con alcances opcionales.

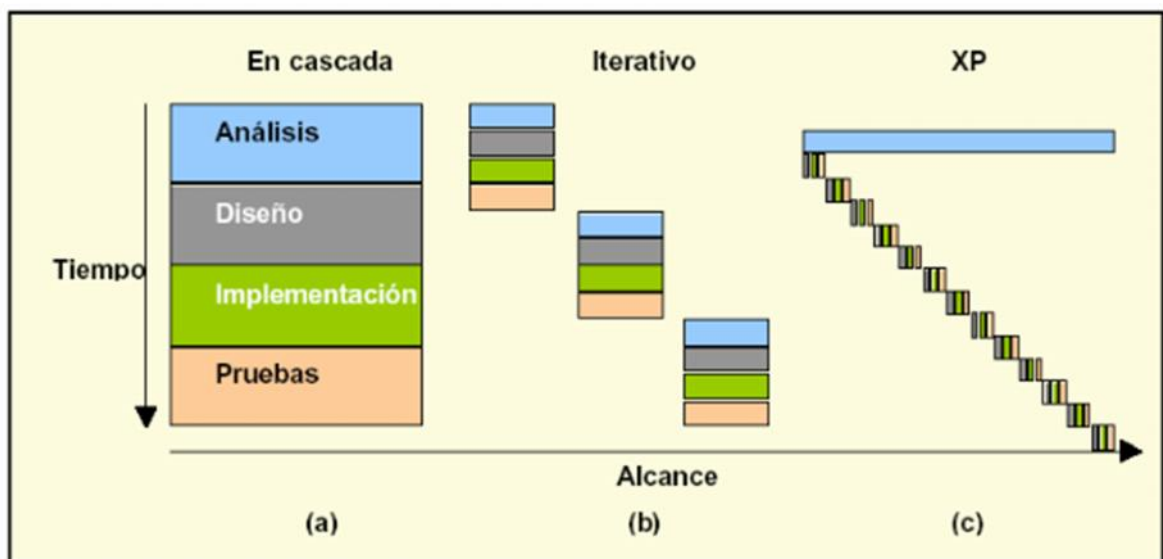
Los ciclos de vida “tradicionales” proponen una clara distinción entre las etapas del proyecto de software, y tienen un plan bien preestablecido acerca del proceso de desarrollo. Asimismo, en todos ellos se parte de especificaciones claras, si no del total del proyecto, por lo menos de una buena parte inicial.

El ciclo de vida de un proyecto XP incluye, al igual que las otras metodologías, entender lo que el cliente necesita, estimar el esfuerzo, crear la solución y entregar el producto final al cliente. Sin embargo, XP propone un ciclo de vida dinámico, donde se admite expresamente que, en muchos casos, los clientes no son capaces de especificar sus requerimientos al comienzo de un proyecto.

Por esto, se trata de realizar ciclos de desarrollo cortos (llamados iteraciones), con entregables funcionales al finalizar cada ciclo. En cada iteración se realiza un ciclo completo de análisis, diseño, desarrollo y pruebas, pero utilizando un conjunto de reglas y prácticas que caracterizan a XP.

Típicamente un proyecto con XP lleva 10 a 15 ciclos o iteraciones. La siguiente figura esquematiza los ciclos de desarrollo en cascada e iterativos tradicionales (por ejemplo, incremental o espiral), comparados con el de XP.

Figura N° 1:
Ciclo de vida de metodología XP



Fuente: Reglas y prácticas en eXtreme Programming, Ing. José Joskowicz

Si bien el ciclo de vida de un proyecto XP es muy dinámico, se puede separar en fases.

Fase de exploración

Es la fase en la que se define el alcance general del proyecto. En esta fase, el cliente define lo que necesita mediante la redacción de sencillas “historias de usuarios”. Los programadores estiman los tiempos de desarrollo en base a esta información. Debe quedar claro que las estimaciones realizadas en esta fase son primarias (ya que estarán basadas en datos de muy alto nivel), y podrían variar cuando se analicen más en detalle en cada iteración.

Esta fase dura típicamente un par de semanas, y el resultado es una visión general del sistema, y un plazo total estimado.

Fase de planificación

La planificación es una fase corta, en la que el cliente, los gerentes y el grupo de desarrolladores acuerdan el orden en que deberán implementarse las historias de usuario, y, asociadas a éstas, las entregas. Típicamente esta fase consiste en una o varias reuniones grupales de planificación. El resultado de esta fase es un Plan de Entregas, o *“Release Plan”*.

Fase de iteraciones

Esta es la fase principal en el ciclo de desarrollo de XP. Las funcionalidades son desarrolladas en esta fase, generando al final de cada una un entregable funcional que implementa las historias de usuario asignadas a la iteración. Como las historias de usuario no tienen suficiente detalle como para permitir su análisis y desarrollo, al principio de cada iteración se realizan las tareas necesarias de análisis, recabando con el cliente todos los datos que sean necesarios. El cliente, por lo tanto, también debe participar activamente durante esta fase del ciclo.

Las iteraciones son también utilizadas para medir el progreso del proyecto. Una iteración terminada sin errores es una medida clara de avance.

Fase de puesta en producción

Si bien al final de cada iteración se entregan módulos funcionales y sin errores, puede ser deseable por parte del cliente no poner el sistema en producción hasta tanto no se tenga la funcionalidad completa.

En esta fase no se realizan más desarrollos funcionales, pero pueden ser necesarias tareas de ajuste (*“fine tuning”*).

A continuación mostramos algunos pantallazos del aplicativo que desarrollamos aplicando la metodología XP la misma que solo contiene una interfaz principal de inicio y las interfaces para el buscador simple y otro para el buscador semántico.

Figura N° 2:
Interfaz de inicio del buscador

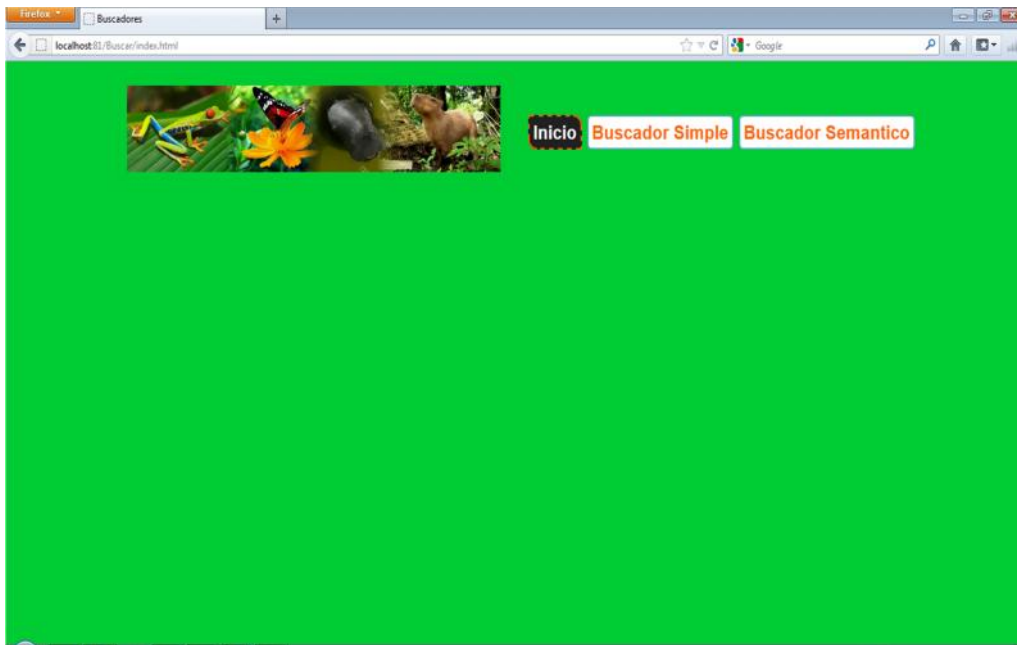


Figura N° 3:
Interfaz de búsqueda con buscador simple

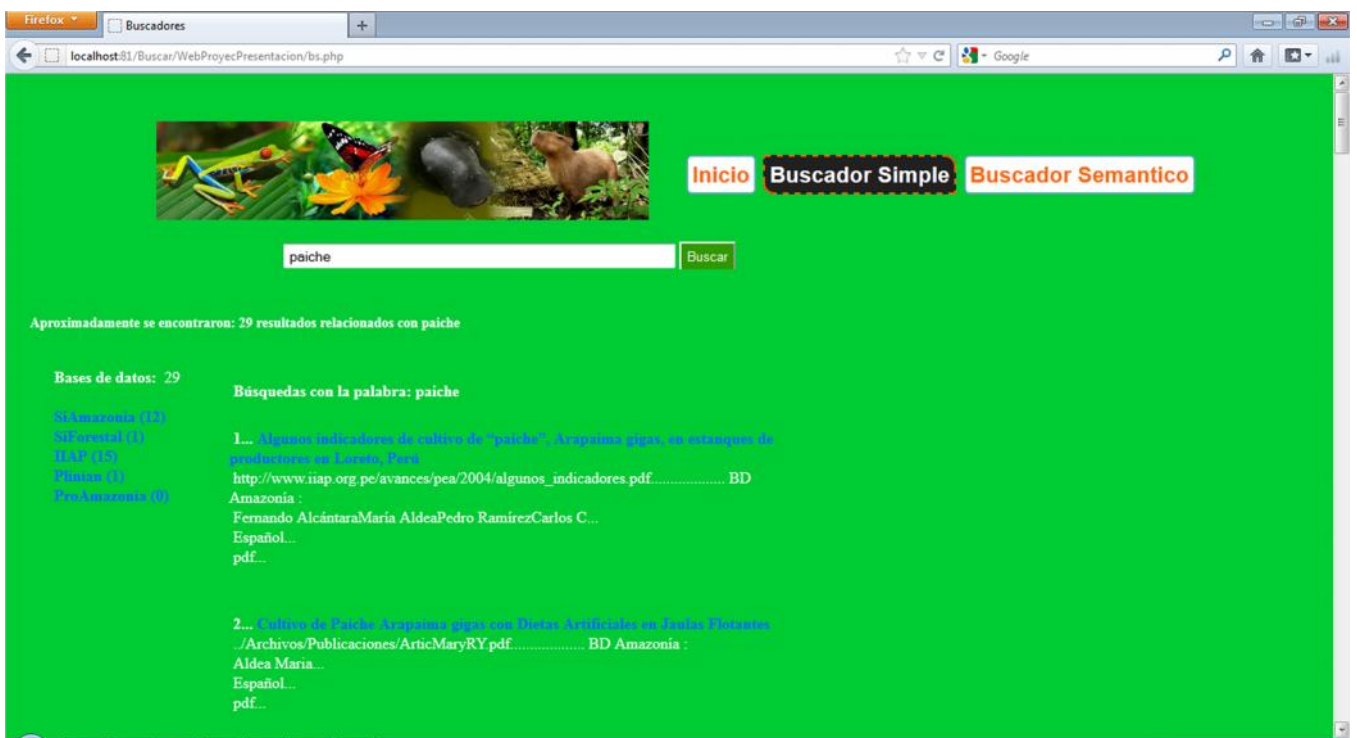
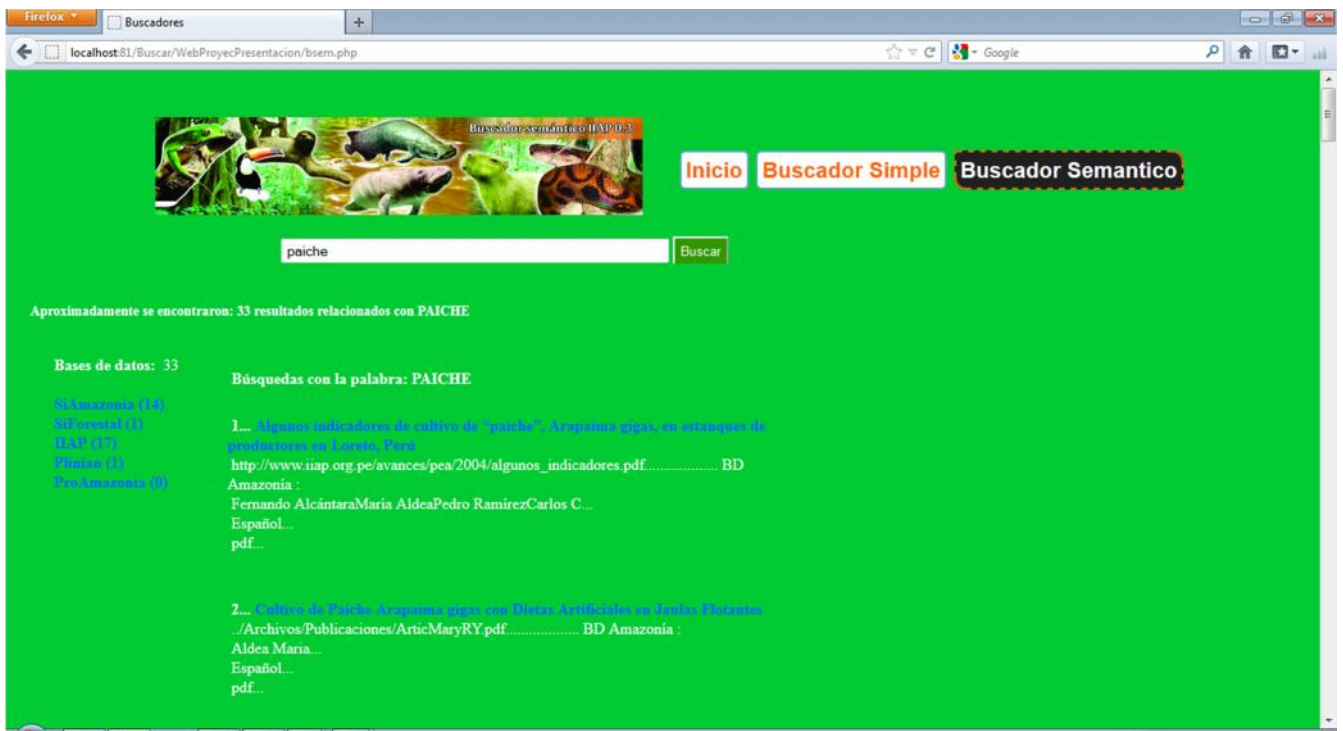


Figura N° 4:
Interfaz de búsqueda con buscador semántico



La figura N° 3 muestra los resultados de búsqueda de la palabra “paiche” (nombre común de la especie) utilizando el buscador simple, obteniendo hasta 29 resultados y la figura N° 4 muestra los resultados de la misma palabra pero se obtiene hasta 33 resultados es decir se encontró 4 resultados más que podrían ser de mucha utilidad para el usuario.

Como se muestra es un aplicativo sencillo y que no necesita mucho tiempo para su desarrollo e implementación siendo el motivo principal para la utilización de la metodología XP por ser ágil.

3.6.2. Método para creación de ontología

a. Dominio y alcance de la ontología

Cuestiones básicas

- Cuál es el dominio que cubre la ontología
- Para qué se usará la ontología
- Para qué tipo de cuestiones
- Quién usará o mantendrá la ontología

Cuestionario de competencia.

Una forma de determinar el alcance de la ontología es bosquejar un cuestionario de competencia, que deberá ser de responder (test):

- La ontología deberá permitir responder al cuestionario.
- Permite saber si es preciso mejorar el nivel de detalle de cierta parte de la ontología.

b. Plantéese reutilizar una ontología existente

- Facilita la integración con otros sistemas.
- Es posible importar y exportar de diversos formalismos (automáticamente o a mano)
- Hay ontologías públicas: Ontolingua, DAML, UNSPSC, RosettaNet, DMOZ, CyC..

c. Enumere términos importantes

Escriba la lista de términos que se usarían para construir sentencias sobre el tema.

- Qué términos nos gustaría usar
- Qué propiedades tienen estos términos.
- Qué podemos decir sobre estos términos.

Ejemplo:

- Vino, uva, bodega, ubicación, color, cuerpo...
- Pescado, carne roja,

No debe importar el solapamiento de términos, si son relaciones o términos o si son clases o slots.

d. Defina las clases y la jerarquía

Aproximaciones:

- Top-down: Primero conceptos generales, después las especializaciones.

Ejemplo: Primero “vino”, “comida”. Después “vino tinto”, “vino rosado”, ... Después “Burdeos”, “Cabernet Sauvignon”, ...

- Bottom-up: Primero las clases más específicas. Las más generales se construyen por agrupamiento.
- Híbrida: de las dos anteriores.

Ninguna de ellas es necesariamente mejor

Habitualmente los conceptos de nivel intermedio son los más manejables:

- No son demasiado abstractos
- Presentan alguna estructura. No así los de bajo nivel.
Depende de la experiencia del diseñador.
De la lista de términos se seleccionan aquellos que describen objetos con existencia independiente.

e. Definir las propiedades de las clases – slots

La mayoría de “términos” restantes alude a propiedades de estas clases

Ejemplo: el color, sabor, cuerpo de un vino.

La ubicación de una bodega.

Debemos determinar a qué clase describe cada término

- Estas propiedades reflejan slots ligados a las clases.
- Ojo, porque esta es la aproximación de Prótegé.

Tipos de slots

- Intrínsecos
Ej: sabor, color del vino,...
- Extrínsecos
Ej: el nombre del vino, el área,...
- Partes
Ej: Los diferentes ingredientes de una comida,...
- Relaciones entre instancias de la clase y otros elementos.
Ej: embotellador, la uva,...

Las subclases heredan los slots.

Se puede aumentar su número.

Ej: en el caso del vino tinto: nivel de tanino.

El slot debe agregarse en el nivel más externo posible.

f. Definir los facetas de cada slot

Los facetas describen el

- Tipo del valor (string, número, booleano, enumerado, instancia de otra clase)
Ej: produce: es de tipo instancia de vino.
- La cardinalidad de la faceta, Simple, múltiple (acotado, libre)
Ej: bodega tiene cardinalidad múltiple.
- Rango del slot: clases permitidas para los slots de tipo instancia
Ej: el rango de produce es vino
- Dominio del slot: clases descritas por el slot

Ej: el dominio de produce es “bodega”

Tipos de valor de los slot:

- String (cadena de caracteres)
- Number (numéricos)
- Boolean (si/no)
- Enumerated (una lista específica de valores admitidos: fuerte, moderado y delicado)
- Instance (admiten la definición de relaciones entre individuos). Por ejemplo, el slot produce de la clase Establecimiento vinícola puede tener instancias de la clase Vino como sus valores.

El dominio|rango deberá ser lo más general posible.

Pero sin pasarse: todas las clases del dominio de un slot se describirán con él y las instancias son potenciales “usuarios” del slot.

Si un rango|dominio incluye una clase y sus subclases, elimine las subclases.

Ej: el rango de produce es “vino”, no “vino tinto”

Si un rango|dominio contiene todas las subclases de una superclase, póngase solo la superclase.

Ej: el dominio rango de produce es “vino”, no “vino tinto”, “rosado”, ...

Si un rango|dominio incluye casi todas las subclases de una superclase, considere afectar a toda la superclase.

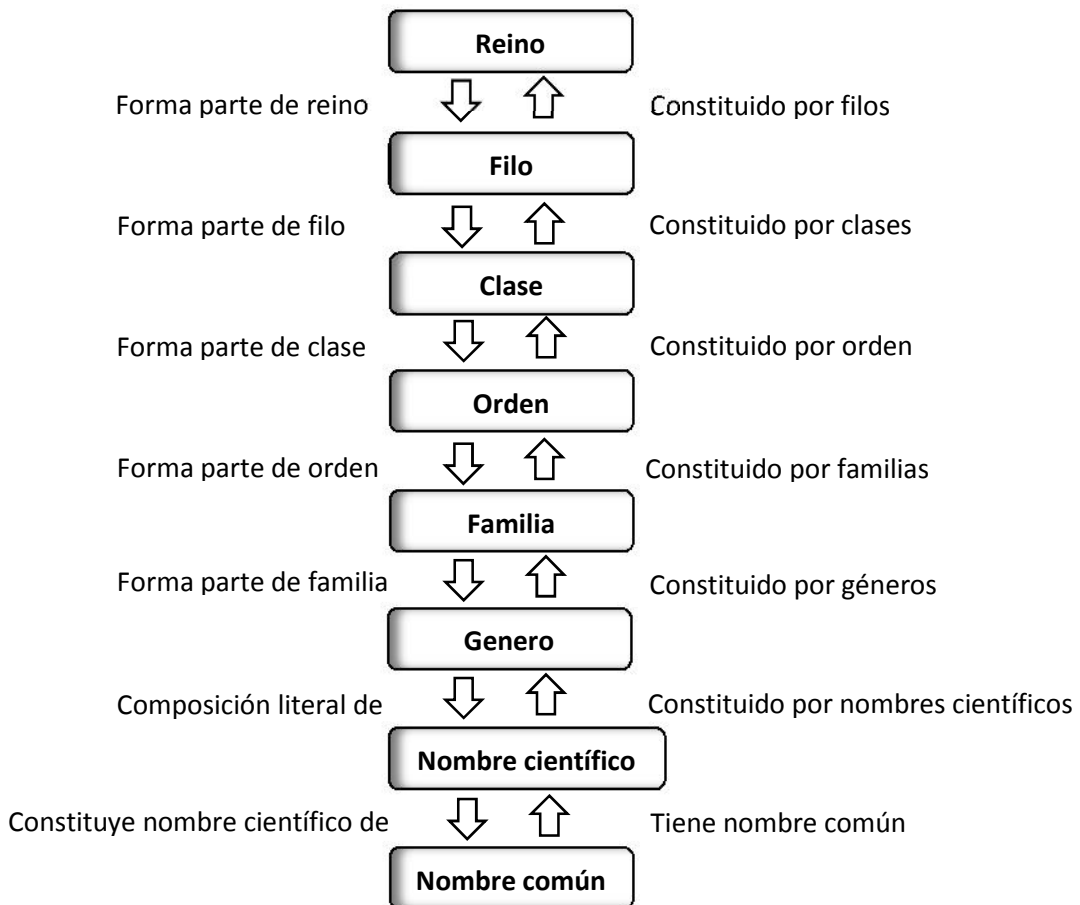
g. Cree las instancias

Proceso:

1. Elegir la clase
2. Crear una instancia de la clase
3. Completar los slots

Finalmente nuestra ontología está conformada de la siguiente manera:

Figura N° 5:
Ontología básica sobre biodiversidad



Fuente: Elaboración propia

3.7. Herramientas para desarrollo de software

3.7.1. Lenguaje de programación PHP

PHP corresponde a las iniciales de personal home page tolos (herramientas para páginas iniciales personales). Es un lenguaje de programación tipo script para entornos web con unas funciones muy semejantes a las de ASP y JSP, utilizado, sobre todo, en servidores Linux para personalizar la información enviada a los usuarios que acceden a un sitio web.

Desde un punto de vista técnico, es un lenguaje interpretado de alto nivel, similar en construcciones léxicas y sintácticas a C, C++, Java y Perl, por lo que a quienes ya conozcan estos lenguajes les resultará muy fácil comenzar a escribir código PHP. PHP es un lenguaje incrustado (embedded) en páginas HTML, es decir, es un lenguaje de programación que se introduce dentro de las páginas HTML. El código PHP se interpreta en el lado del servidor de web, desde donde se genera la página HTML solicitada antes de llevar a cabo su transmisión al navegador.

De esta forma, podemos programar aplicaciones asociadas al servidor de web, aumentando, así, la funcionalidad de dicho servidor y convirtiéndolo en un sistema de desarrollo de aplicaciones cliente/servidor mucho más completo. Su principal objetivo es hacer que desarrolladores de aplicaciones basadas en la web puedan escribir páginas que se generan dinámicamente de una forma sencilla y rápida.

En cuanto a la tecnología del intérprete de PHP, la versión 3 ya era tan rápida como los intérpretes existentes de ASP. Con la versión 4 de PHP, su rendimiento y prestaciones mejoraron todavía más: el intérprete (Zend) era hasta 12 veces más rápido que el de la versión 3; se modularizó todo el diseño interno; se perfeccionó su integración con otros servidores HTTP como el IIS de Microsoft, y se encaró hacia la programación orientada a objetos (Programación OO). Con la versión 5, se ha rediseñado completamente el motor Zend, para crear un lenguaje completamente OO, agilizando más aún su funcionamiento, y extrayendo la compatibilidad con MySQL en un módulo externo (por cuestiones de licencia con MySQL, este SGBD ha dejado de ser "la base de datos" de PHP, para ser una más de las que PHP puede tratar).

PHP proporciona, por tanto, una gran facilidad para acceder a diferentes tipos de bases de datos como Oracle, Sybase, MySQL, PostgreSQL, Adabas, etc. De hecho, es bastante sencillo portar una aplicación escrita con PHP para MySQL a cualquier otro servidor de base de datos, ya que las funciones de acceso que ofrece PHP son, en muchos casos, de sintaxis compartida.

3.7.2. MySql

MySQL es un sistema de gestión de bases de datos (SGBD) SQL que en algunos aspectos es aproximadamente tan potente como Oracle. Cabe mencionar que a mediados del año 2009, Oracle, ha adquirido MySQL.

Sus principales objetivos han sido la velocidad y la robustez. Es un SGBD sencillo y rápido que se adapta perfectamente a entornos en los que el volumen de datos sea del orden de megabytes (en la documentación se habla de su uso con bases de datos de 50 millones de registros). En la versión 5 de MySQL ha incluido el control de transacciones, procedimientos almacenados y *triggers*, por lo que ha rellenado el gran hueco que lo diferenciaba de grandes SGBD como Oracle. Si bien existe la posibilidad de comprar su soporte, Oracle sigue teniendo más aceptación en el mundo empresarial.

En el último *benchmark* realizado de SGBD, se ha visto un gran crecimiento en la potencia de MySQL, frente a sus competidores.

En estos últimos años destacadas compañías de software han desarrollado aplicaciones SQL de uso libre y con código fuente (*open source*). En el mundo de GNU/Linux MySQL es, junto a Postgres (<http://www.postgresql.org/>), uno de los SGBD más populares.

MySQL también puede verse como un conjunto de aplicaciones o *pluggins* funcionando en conjunto.

Existen en su versión actual distintos motores de almacenamiento de datos, entre los que destacan MyISAM (permite índices por cadenas completas) y InnoDB (que permite el uso de transacciones) o la incorporación de *buffers* en memoria que permiten agilizar la respuesta de sus resultados.

MySQL se encuentra, igual que PHP, en fase de pleno desarrollo; se están publicando con regularidad nuevas versiones del sistema, así como herramientas que son básicas en cualquier SGBD actual:

- Dispositivo JDBC para acceder desde Java.
- Dispositivo ODBC para acceder utilizando la API ODBC.
- API de programación para C, Perl, C++, Python y TCL.
- Acceso desde PHP.
- Entornos visuales de gestión de la base de datos.
- Control de acceso basado en una base de datos de administración.

3.7.3. Lenguaje RDF

Es la Infraestructura para la Descripción de Recursos (Resource Description Framework). Es un formato universal para representar datos en la Web, que va a permitir intercambiar información a través de diferentes aplicaciones sin que esos datos pierdan significado, lo que facilita la reutilización de los recursos en la Web.

RDF es un modelo de metadatos que dirige por referencia, muchos de los aspectos de codificación que requiere el almacenamiento y transferencia de archivos (tales como internacionalización, conjuntos de caracteres, etc.).

Para estos aspectos, RDF cuenta con el soporte de XML. Es importante también entender que esta sintaxis XML, es sólo una sintaxis posible para RDF y que pueden surgir formas alternativas para representar el mismo modelo de datos

El objetivo general de RDF es definir un mecanismo para describir recursos que no cree ninguna asunción sobre un dominio de aplicación particular, ni defina (a priori) la semántica de algún dominio de aplicación. La definición del mecanismo debe ser neutral con respecto al dominio, sin embargo el mecanismo debe ser adecuado para describir información sobre cualquier dominio.

RDF (Resource Description Framework)

- Permite describir recursos mediante propiedades y valores de propiedades

- Recurso: Cualquier cosa que pueda nombrarse mediante una URI
- Propiedad: Característica o atributo de un recurso
 - Tiene asociada una URI y un significado concreto
 - Puede relacionarse con otras propiedades
- Enunciado: Asocia el valor de una propiedad a un recurso

RDF Schema

- Permite definir una clasificación de recursos en términos de clases, subclases, propiedades, etc.
- Establece restricciones sobre las posibles combinaciones
- Elementos:
 - Resource: Todo son recursos
 - Class: Tipo o categoría
- Las clases pueden ser subclases de otras
- El tipo de un recurso debe ser un elemento de una clase
 - Property: Subconjunto de recursos que son propiedades
- Las propiedades tienen un rango y un dominio

3.7.4. Protégé

- Editor de ontologías y bases de conocimiento gratis y abierto
- Basado en Java
- Soporta Frames, XML Schema, RDF y OWL
- Cuenta con un ambiente “plug-and-play”

<http://protege.stanford.edu/>

Protégé es una herramienta open source desarrollada en Stanford Medical Informatics. Como muchas otras herramientas de modelamiento, Protégé se encuentra separada entre una parte de modelo y una parte de vista.

El modelo de Protégé es el mecanismo de representación interna de ontologías y bases de conocimiento. Los componentes de la vista de Protégé proveen al usuario una interface para visualizar y manipular el modelo (“Getting Started With Protégé”, 2003).

Figura N° 6:

Creación de clases de la ontología básica sobre biodiversidad utilizando Protégé

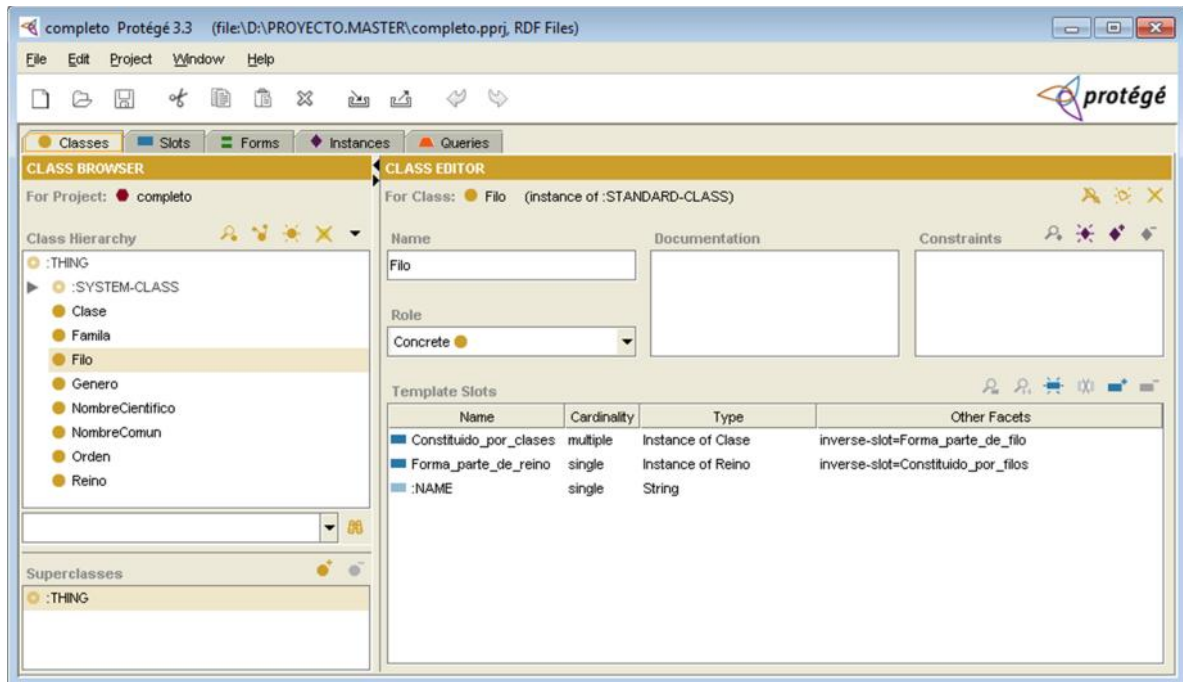
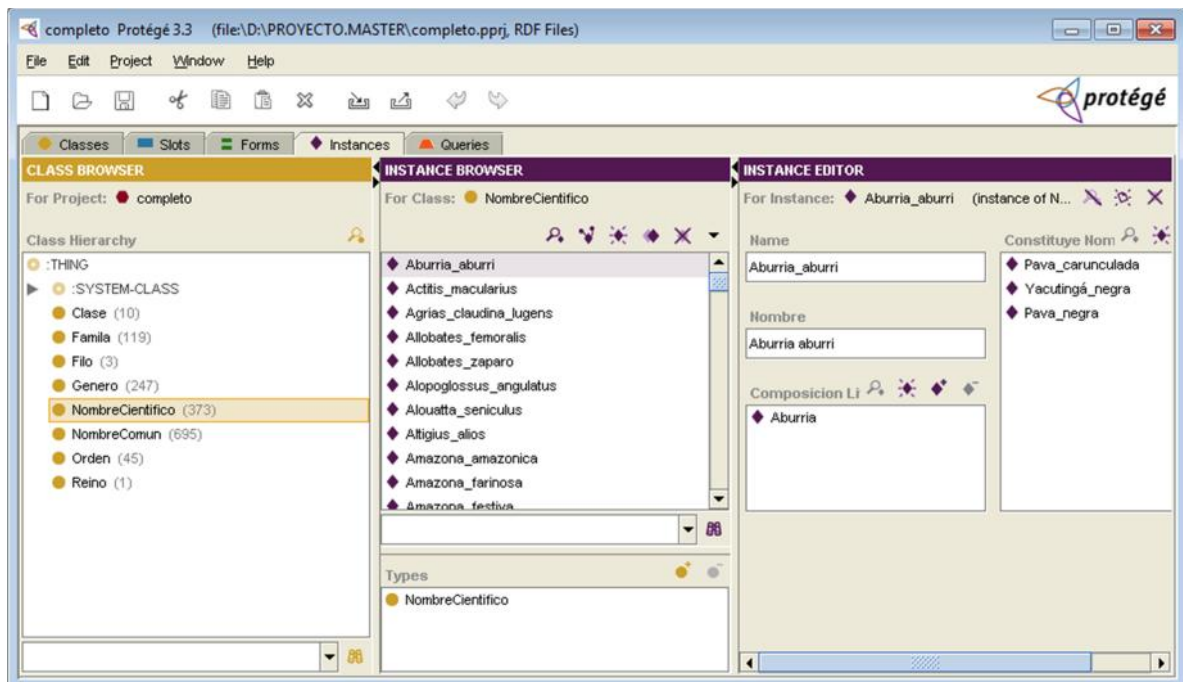


Figura N° 7:

Creación de instancias de la ontología básica sobre biodiversidad utilizando Protégé



El modelo de Protégé está basado en un simple metamodelo comparable con sistemas orientados a objetos y sistemas basados en marcos. Éstos básicamente pueden representar ontologías que consisten en clases, propiedades, características de propiedades e instancias. La ventaja de Protégé es que su metamodelo es una ontología Protégé. La estructura del metamodelo de Protégé habilita fácilmente la extensión y adaptación a otras representaciones.

Usando las vistas de la interface de usuario de Protégé, los diseñadores de ontologías básicamente crean clases y les asignan propiedades. Usando la función *resulting ontologies*, Protégé es capaz de generar automáticamente interfaces de usuario que soporten la creación de instancias. En cada clase de la ontología el sistema crea un formulario con componentes de edición para cada propiedad de la clase.

Protégé actualmente puede ser usado para cargar, editar y salvar ontologías en varios formatos como CLIPS, RDF, XML, UML, OWL. Las ontologías tienen un papel importante en las aplicaciones de web semántica. Las extensiones de Protégé pueden beneficiar desde los servicios generales proveídos de la plataforma núcleo, hasta un mecanismo de mecanismos de plugin.

Con el plugin OWL de Protégé se puede permitir que mucha gente edite la misma ontología al mismo tiempo. Protégé también provee una base de datos altamente escalable, lo que permite que los usuarios creen ontologías con muchas clases (Knublauch, 2005).

El plugin OWL es una extensión compleja que puede ser usada para editar archivos OWL y bases de datos. El plugin incluye una colección de características para los componentes denominados *widgets* y provee acceso a servicios basado en clasificación, chequeo consistente y pruebas.

Protégé OWL es una herramienta que hace posible la descripción de conceptos y provee otras facilidades como por ejemplo: tiene un enriquecido conjunto de operadores (*and*, *or* y la negación); está basado sobre un modelo lógico que garantiza que los conceptos estén bien descritos; se pueden describir conceptos complejos, y el modelo lógico permite usar un razonador que chequea la consistencia descrita en la ontología garantizando que la jerarquía esté correcta.

3.8. Desarrollo de la investigación

En relación con las hipótesis de esta investigación, ya sabemos que se llevó a cabo un estudio cuasiexperimental con pre test/post test, sin grupo de control. El estudio experimental solo se realizó a un grupo de cincuenta alumnos de la facultad de ciencias biológicas de la UNAP y en un solo día (12.Nov.2014 en los ambientes del Instituto de Investigaciones de la Amazonía Peruana – IIAP), durante el cual se enseñó a utilizar dos interfaces de buscadores (un buscador simple y un buscador semántico) con funcionalidades similares a simple vista.

El objetivo de este estudio fue determinar si el modelo semántico sustentado en el uso de buscadores era eficiente para encontrar información sobre biodiversidad. Los resultados deberían reflejar que las búsquedas a través de la modalidad presencial y de pruebas de búsqueda incrementarían la eficiencia del buscador semántico específicamente.

Es preciso indicar que los estudiantes pertenecen al tercer nivel de la carrera de ciencias biológicas, y presentaban cierto nivel de aprendizaje necesario para las pruebas, ya que era muy necesario que tengan los conocimientos básicos sobre especies y sus nombres científicos. El promedio de edades de los alumnos era de 23 años, del total 40% (20) eran del sexo masculino y 60% (30) del sexo femenino.

3.8.1. La estructura del pre test y post test

Es importante señalar que la estructura del pre test y del post test, que integra las preguntas relacionadas a cada uno de los indicadores de las variables independiente y dependiente, también ha sido adaptada en ciertos aspectos acorde con el objetivo general y los objetivos específicos.

La elaboración de dichos cuestionarios se realizó con los parámetros y el formato de elaboración propia teniendo en cuenta una escala evaluativa estándar ya que son muy usados en temas de investigación, siendo el mismo esquema de evaluaciones para el pre test y post test (ver anexo N° 2).

3.8.2. Implementación del cuestionario a los participantes

La temática de las pruebas in situ fue de la siguiente manera: en un espacio de las instalaciones del proyecto BIOINFO del IIAP se pusieron en funcionamiento diez (10) laptops con acceso a la intranet, de manera que puedan conectarse al servidor a través de una ruta de acceso a las interfaces de los buscadores, es decir se puedan

enlazar al link <http://200.60.174.203/buscar/>. Luego se invitó a los alumnos a pasar en grupos de diez para iniciar su intervención el cual debería durar no más de 10 minutos, a los cuales también se orientaba sobre lo que tenían que hacer dentro del link.

Dentro de ello el alumno participante se encontraba con una interfaz principal que mostraba tres “botones”, uno con el nombre de “Inicio” y otros con los nombres de “Buscador simple” y “Buscador semántico”. Al hacer clic en el botón “Buscador simple” o “Buscador semántico” se mostraba una caja de texto y otro botón con el nombre de “Buscar” de manera que pueda realizar su búsqueda sobre las especies antes escogidas de una lista predefinida y otorgadas por nosotros.

Posterior a la observación, interacción y validación realizada por el alumno con los buscadores antes mencionados, ya estaba preparado para poder evaluar ambos prototipos de manera que su evaluación para el “buscador simple” lo tenía que realizar dentro de la estructura del cuestionario de pretest y para el “buscador semántico” lo tenían que realizar dentro del post test. Siendo el final de su intervención.

IV. Resultados

4.1. Evaluación de la implementación y uso del buscador semántico

Al analizar las respuestas de la encuesta del post test, los mismos que están relacionadas con los componentes de la variable independiente, buscador semántico, con sus indicadores: interfaz rápida, secuencia a seguir, tiempo de respuesta y listado de registro, cuyas valoraciones para su evaluación son: muy apropiado - MA (puntuación 5), ligeramente apropiado - LA (puntuación 4), apropiado - A (puntuación 3), ligeramente inapropiado - LI (puntuación 2) y muy inapropiado - MI (puntuación 1); mostramos en la siguiente tabla la situación en que se encuentra la implementación y uso de nuestro buscador semántico, del cual hemos obtenido los siguientes resultados:

Tabla N° 01

Evaluación del uso del buscador semántico realizado por los estudiantes de la Facultad de Ciencias Biológicas de la UNAP - 2014

PREGUNTAS	Post Test - Buscador Semántico										TOTAL	
	MI		LI		A		LA		MA			
	1		2		3		4		5			
	N°	%	N°	%	N°	%	N°	%	N°	%	N°	%
1. INTERFAZ RAPIDA	0	0	0	0	6	12	14	28	30	60	224	25.9
2. SECUENCIA A SEGUIR	0	0	1	2	8	16	16	32	25	50	215	24.9
3. TIEMPO DE RESPUESTA	0	0	1	2	5	10	17	34	27	54	220	25.5
4. LISTADO DE REGISTROS	0	0	3	6	10	20	16	32	21	42	205	23.7
Promedio	0	0	1	3	7	15	16	32	26	52	864	100

Fuente: Elaboración propia

Del promedio de 100% de la valoración asignada a los indicadores, el 52% del valor asignado al buscador semántico es muy apropiado (MA), y también se observa que el valor de muy apropiado es superior para los cuatro (4) indicadores en relación a los demás valores, pero predominando dentro de ello la valoración que se dio al indicador de *interfaz rápida* con un 60% de aceptación. Asimismo, se observar que el 32% del valor asignado

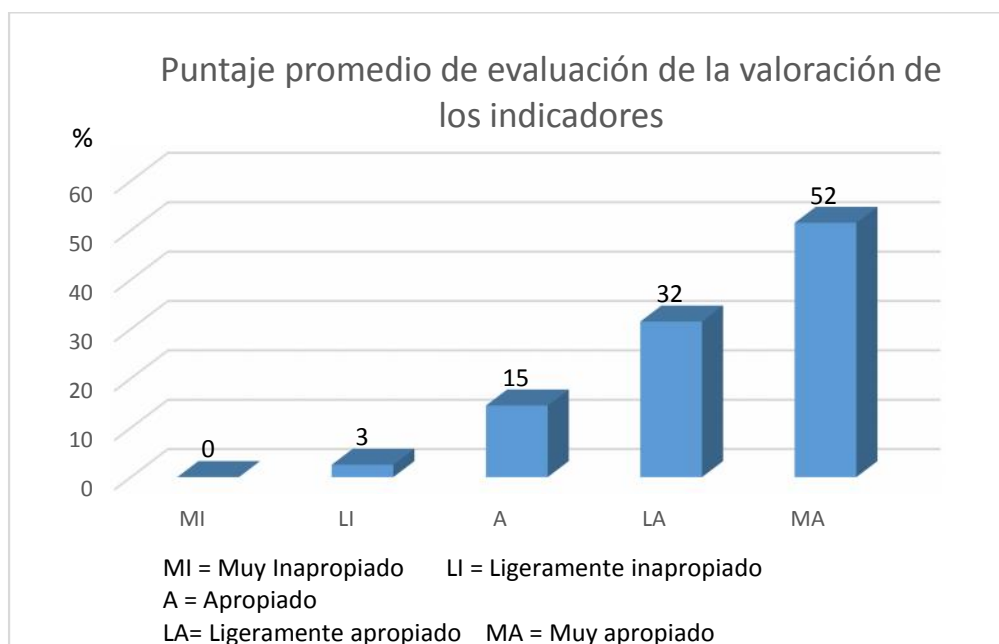
considera que el buscador semántico es ligeramente apropiado (LA), predominando dentro de ello el indicador de *tiempo de respuesta* con un 34% de aceptación. También podemos observar que los estudiantes en ningún momento consideran al buscador semántico como muy inapropiado (MI) y que solamente el 3% del valor asignado lo considera como ligeramente inapropiado (LI).

Asimismo, del promedio de 100% de la evaluación de los indicadores de la implementación y uso del buscador semántico en relación a sus cuatro (4) valores, predomina el indicador de *interfaz rápida* con el 25.9% de la evaluación promedio asignada, estando muy cerca el indicador de *tiempo respuesta* con un 25.5% de la evaluación promedio asignado.

Estos resultados se evidencian en los siguientes gráficos.

Gráfico N° 01:

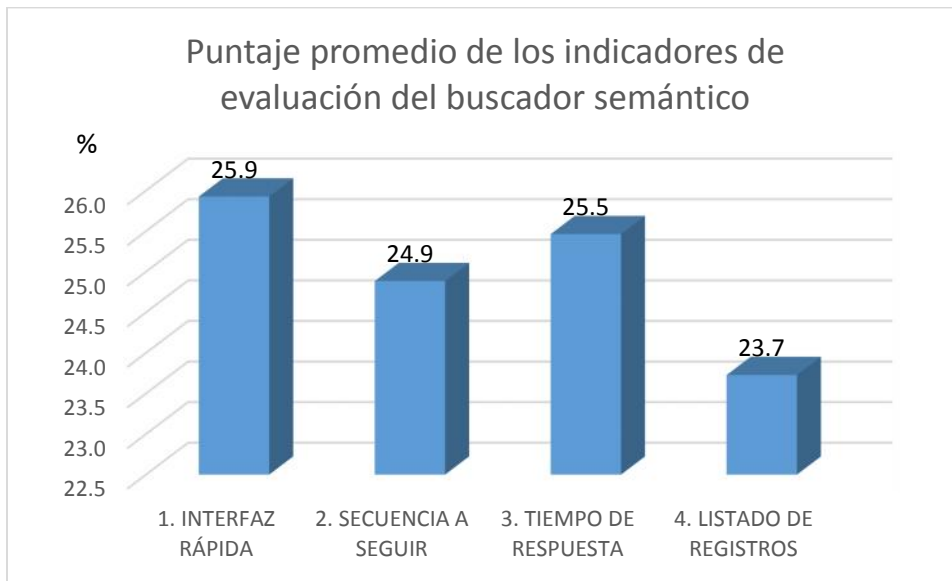
Puntaje promedio de la valoración de indicadores del uso del buscador semántico



Fuente: Elaboración propia

Gráfico N° 02:

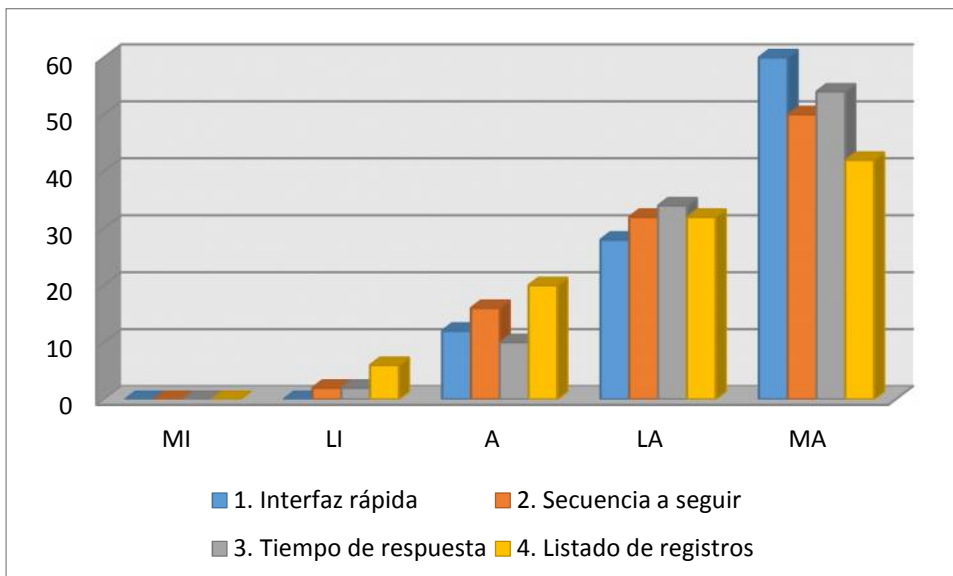
Puntaje promedio de los indicadores de evaluación del uso del buscador semántico



Fuente: Elaboración propia

Gráfico N° 03:

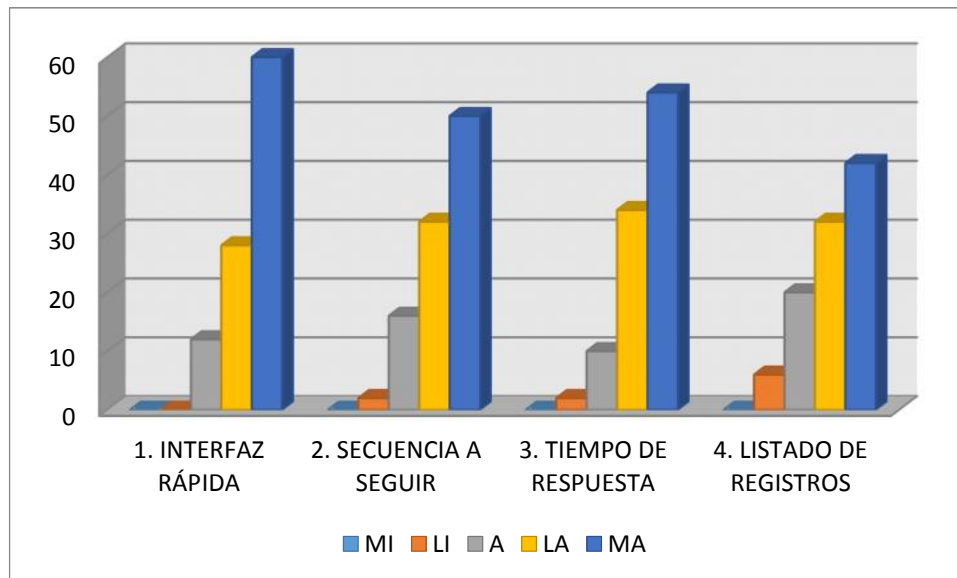
Evaluación promedio de la valoración de los indicadores del uso del buscador semántico



Fuente: Elaboración propia

Gráfico N° 04:

Evaluación promedio de los indicadores del uso del buscador semántico



Fuente: Elaboración propia

4.2. Evaluación de la eficiencia en los resultados de búsqueda de información sobre biodiversidad amazónica

Posterior al procesamiento de las encuestas del Buscador Simple (pre test) y Buscador Semántico (post test), se ha elaborado las tabla N°02 y tabla N°03, y se ha realizado una comparación entre ambos (teniendo en cuenta los mismos cuatro indicadores de medición).

Tabla N°02
Evaluación pre test de la eficiencia de búsqueda de información sobre biodiversidad realizada al buscador simple, por los estudiantes de la Facultad de Ciencias Biológicas de la UNAP – 2014

PREGUNTAS	Pre test - Buscador Simple										TOTAL		PROM
	MI		LI		A		LA		MA				
	1	2	3	4	5	N°	%	N°	%	N°	%	N°	
1. TIEMPO TOTAL DE BÚSQUEDA	6	12	10	20	14	28	10	20	10	20	158	25,4	3.16
2. CANTIDAD DE PASOS A SEGUIR	4	8	12	24	12	24	11	22	11	22	163	26,2	3.26
3. CANTIDAD DE REGISTROS	5	10	11	22	22	44	7	14	5	10	146	23,4	2.92
4. UTILIDAD DE LA INFORMACIÓN	7	14	10	20	13	26	10	20	10	20	156	25,0	3.12
Promedio	5,5	11	11	22	15	30,5	9,5	19	9	18	623	100	

Fuente: Elaboración propia

Tabla N°03

Evaluación post test de la eficiencia de búsqueda de información sobre biodiversidad realizada al buscador semántico, por los estudiantes de la Facultad de Ciencias Biológicas de la UNAP – 2014

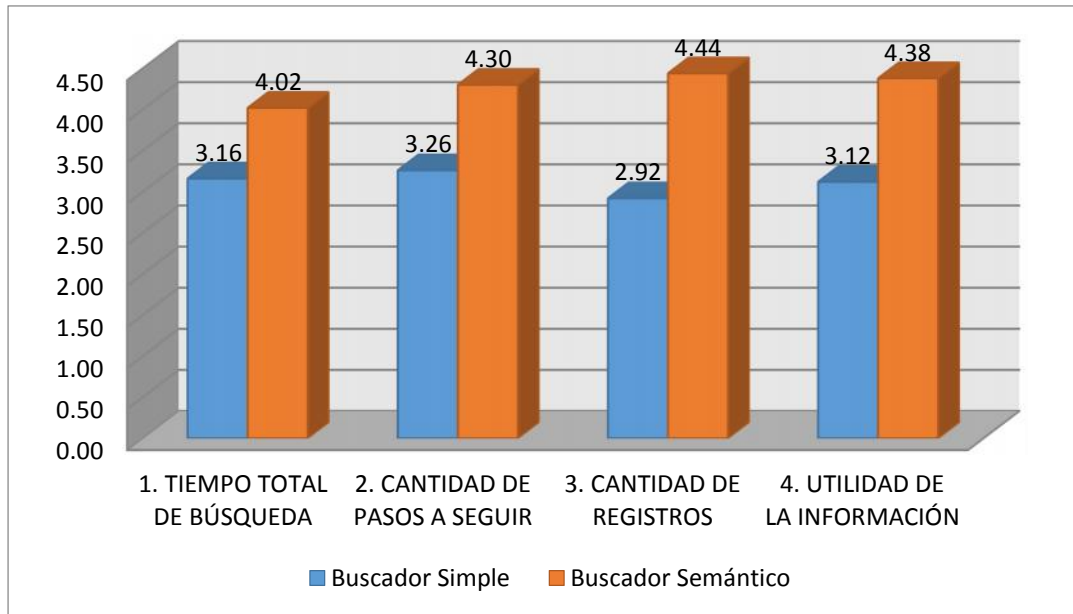
PREGUNTAS	Post Test - Buscador Semántico										TOTAL		PROM
	MI		LI		A		LA		MA				
	1	2	3	4	5	N°	%	N°	%	N°	%	N°	
1. TIEMPO TOTAL DE BÚSQUEDA	0	0	1	2	14	28	18	36	17	34	201	23,5	4.02
2. CANTIDAD DE PASOS A SEGUIR	0	0	0	0	10	20	15	30	25	50	215	25,1	4.30
3. CANTIDAD DE REGISTROS	0	0	0	0	8	16	12	24	30	60	222	25,9	4.44
4. UTILIDAD DE LA INFORMACIÓN	0	0	0	0	5	10	21	42	24	48	219	25,6	4.38
Promedio	0	0	0,25	0,5	9,25	18,5	16,5	33	24	48	857	100	

Fuente: Elaboración propia

Encontramos que en los cuatros indicadores estudiados los resultados son favorables para el Buscador Semántico, es decir todos los valores promedio de la variable estudiada son superiores a los valores del Buscador Simple. Además, respecto a las diversas comparaciones de los indicadores entre ambos buscadores podemos destacar que el más alto promedio comparativo está relacionado a la **cantidad de registros** obtenidos, e incluso podemos concluir que la principal diferencia entre ambos buscadores se da en la cantidad de registros obtenidos a favor del Buscador Semántico. Por otro lado el indicador que menos diferencia presenta está relaciona al **tiempo total de búsqueda**, donde la diferencia es menor respecto a los otros indicadores. Ello además se puede evidenciar si revisamos el gráfico N° 05.

Gráfico N° 05:

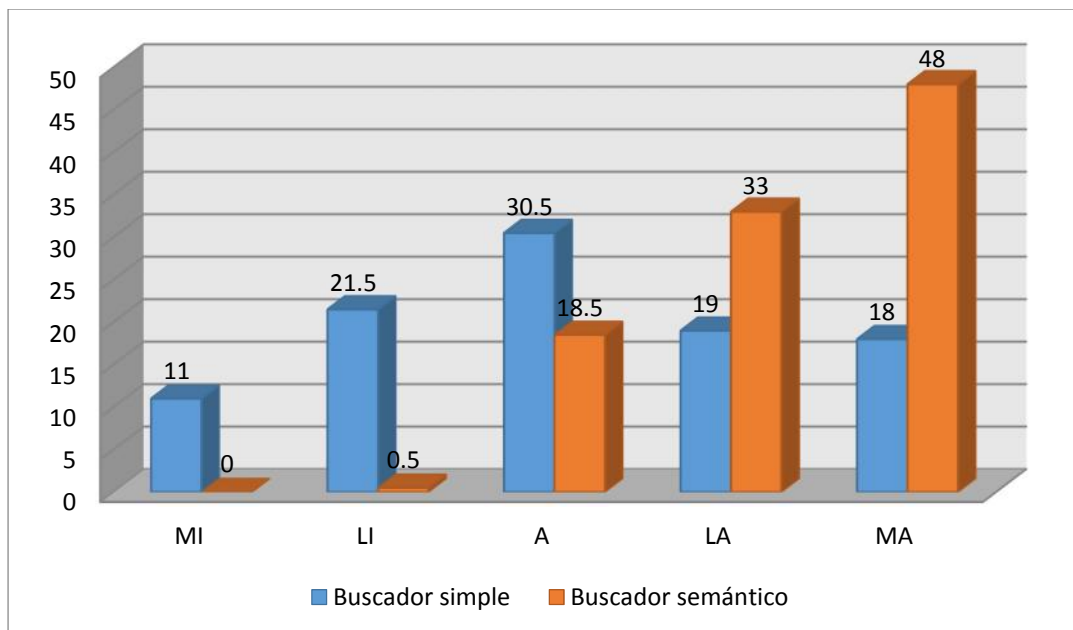
Comparación entre los promedios de los indicadores de eficiencia del buscador simple y buscador semántico.



Fuente: Elaboración propia

Gráfico N° 06:

Comparación entre los promedios de la valoración de los indicadores de eficiencia del buscador simple y buscador semántico.



Fuente: Elaboración propia

4.3. Tamaño del efecto en los resultados de las búsquedas de información sobre biodiversidad amazónica al aplicar el buscador semántico

Para determinar el tamaño del efecto de la variable dependiente: eficiencia en los resultados de las búsquedas de información sobre biodiversidad amazónica, se calculó utilizando la ecuación del efecto determinado por Flores et al (2014) (ver ecuación N° 4.3.3.1) el mismo que utiliza el resultado del t student calculado encontrado en la contrastación de hipótesis realizada como parte de la investigación y que se detalla a continuación.

4.3.1. Contrastación de la hipótesis

La contrastación de la hipótesis se realizó con el método propuesto en la metodología pre test – post test que nos permite aceptar o rechazar la hipótesis. Para esto se realizó una prueba por cada indicador las cuales se detallan en las tablas N° 04 y N° 05 de pre test y post test respectivamente.

Tabla N° 04
Resultados del pre test o buscador simple

PREGUNTAS	Pre test – Buscador Simple										TOTAL		PROM	
	MI		LI		A		LA		MA					
	1	2	3	4	5	N°	%	N°	%	N°	%	N°		%
1. INTERFAZ RÁPIDA	6	12	14	28	6	12	10	20	14	28	162	13.1	3.24	
2. SECUENCIA A SEGUIR	5	4	9	18	16	32	10	20	10	20	161	13.0	3.22	
3. TIEMPO DE RESPUESTA	10	20	10	20	10	20	8	16	12	24	152	12.3	3.04	
4. LISTADO DE REGISTROS	4	8	20	40	12	24	8	16	6	12	142	11.5	2.84	
5. TIEMPO TOTAL DE BÚSQUEDA	6	12	10	20	14	28	10	20	10	20	158	12.7	3.16	
6. CANTIDAD DE PASOS A SEGUIR	4	8	12	24	12	24	11	22	11	22	163	13.1	3.26	
7. CANTIDAD DE REGISTROS	5	10	11	22	22	44	7	14	5	10	146	11.8	2.92	
8. UTILIDAD DE LA INFORMACIÓN	7	14	10	20	13	26	10	20	10	20	156	12.6	3.12	
Promedio	5.9	11	12	24	13	26	9.3	19	9.8	20	1240	100		

Fuente: Elaboración propia

Tabla N° 05
Resultados del post test o buscador semántico

PREGUNTAS	Post Test – Buscador Semántico												TOTAL	PROM
	MI		LI		A		LA		MA					
	1		2		3		4		5					
	N°	%	N°	%	N°	%	N°	%	N°	%	N°	%		
1. INTERFAZ RÁPIDA	0	0	0	0	6	12	14	28	30	60	224	13.0	4.48	
2. SECUENCIA A SEGUIR	0	0	1	2	8	16	16	32	25	50	215	12.5	4.30	
3. TIEMPO DE RESPUESTA	0	0	1	2	5	10	17	34	27	54	220	12.8	4.40	
4. LISTADO DE REGISTROS	0	0	3	6	10	20	16	32	21	42	205	11.9	4.10	
5. TIEMPO TOTAL DE BÚSQUEDA	0	0	1	2	14	28	18	36	17	34	201	11.7	4.02	
6. CANTIDAD DE PASOS A SEGUIR	0	0	0	0	10	20	15	30	25	50	215	12.5	4.30	
7. CANTIDAD DE REGISTROS	0	0	0	0	8	16	12	24	30	60	222	12.9	4.44	
8. UTILIDAD DE LA INFORMACIÓN	0	0	0	0	5	10	21	42	24	48	219	12.7	4.38	
Promedio	0	0	0.8	1.5	8.3	17	16	32	25	49.8	1721	100		

Fuente: Elaboración propia

Tabla N°06
Presentación del calificativo

Abreviatura	Descripción
MI	Muy inapropiado
LI	Ligeramente inapropiado
A	Apropiado
LA	Ligeramente apropiado
MA	Muy apropiado

Fuente: Elaboración propia

4.3.2. Prueba t student

Tabla N°07
Definición de variables

<u>Abreviatura</u>	<u>Descripción</u>
PA_i	Puntuacion del Sistema actual
PP_i	Puntuacion del Sistema Propuesto
D_i	$(PA_i - PP_i)$
D_i^2	$(PA_i - PP_i)^2$

Fuente: Elaboración propia

Tabla N°08
Contrastación entre pre test y post test

PREGUNTAS	PA_i	PP_i	D_i	D_i^2
1. INTERFAZ RÁPIDA	3.24	4.48	-1.24	1.5376
2. SECUENCIA A SEGUIR	3.22	4.3	-1.08	1.1664
3. TIEMPO DE RESPUESTA	3.04	4.4	-1.36	1.8496
4. LISTADO DE REGISTROS	2.84	4.1	-1.26	1.5876
5. TIEMPO TOTAL DE BÚSQUEDA	3.16	4.02	-0.86	0.7396
6. CANTIDAD DE PASOS A SEGUIR	3.26	4.3	-1.04	1.0816
7. CANTIDAD DE REGISTROS	2.92	4.44	-1.52	2.3104
8. UTILIDAD DE LA INFORMACIÓN	3.12	4.38	-1.26	1.5876
SUMATORIA	24.8	34.42	-9.62	11.8604

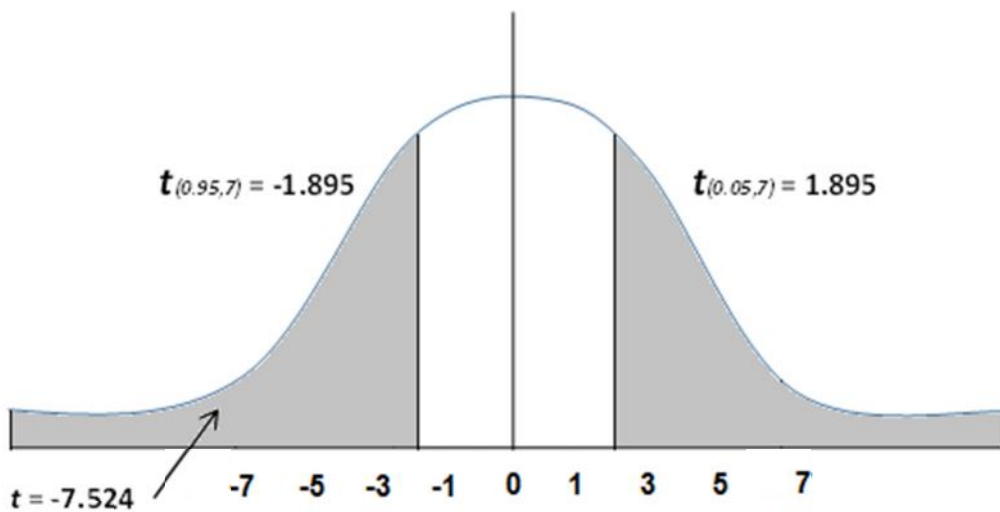
Fuente: Elaboración propia

Tabla N°09
Resultados obtenidos

Diferencia promedio: $D = \sum_{i=0}^n Di / 8$	-1.2025
Desviacion estándar: $S = (8) \left(\sum_{i=1}^n Di^2 \right) - \left(\sum_{i=1}^n Di \right)^2 / (8)(7)$	0.04176429
Varianza: $V = \sqrt{S}$	0.20436312
T calculado: $T_c = ((D)(\sqrt{8}))/\sqrt{V}$	-7.524

Fuente: Elaboración propia

Figura N° 8:
Distribución t student



Puesto que: $T_c = -7.524$ (t calculado) es menor que $T_\alpha = -1.895$ (tabular), con $gl = 7$ (grado de libertad) estando este valor dentro de la región de rechazo se concluye que $T_c - T_\alpha < 0$, se rechaza la hipótesis nula (H_0) y la hipótesis alterna es aceptada (H_a), por lo tanto se prueba la validez de la hipótesis con un nivel de error de 5% siendo: La implementación de un buscador semántico incrementa significativamente la eficiencia en los resultados de las búsquedas sobre

biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014.

4.3.3. Cálculo del tamaño del efecto *TE*

Una vez que calculamos el resultado de *t* calculado (T_c) ahora podemos determinar el tamaño del efecto de la variable dependiente con la siguiente ecuación:

$$TE = \sqrt{\frac{(T_c)^2}{(T_c)^2 + gl}} \quad (4.1)$$

Dónde:

$n = 50$ alumnos

$T_c = T$ calculado

$gl =$ grados de libertad ($n-1$)

Entonces:

$$TE = \sqrt{\frac{(-7.524)^2}{(-7.524)^2 + 49}} = 0.7321$$

Cuando se trabaja experimentalmente se intenta comprobar o poner a prueba la eficiencia de tratamientos.

Después de calcular el tamaño del efecto que es de 0.7321 (73.21%) , significa que la eficiencia en los resultados de las búsquedas de información sobre biodiversidad amazónica al utilizar el buscador simple y semántico respectivamente en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014, tiene un tamaño de efecto grande.

V. Discusión

- En consideración a los indicadores utilizados para evaluar el buscador semántico, se ha comprobado que podemos confiar en los resultados que nos devuelve y muestra ya que al efectuarse la búsqueda sobre algún término de biodiversidad amazónica, la valoración realizada obtuvo un 52% de aceptación para el valor de *muy apropiado*, es decir gracias a la implementación que ha respondido con éxito y por el buen uso que le dieron los estudiantes de ciencias biológicas de la UNAP, esperando que en el futuro puedan seguir usando el buscador semántico para complementar sus investigaciones y trabajos académicos profesional; sin embargo Criado (2009), nos indica en su tesis que existen implementaciones tanto para buscadores semántico genéricos o de propósito general y buscadores semánticos especializados para una actividad en concreto, de acuerdo a lo indicado nuestro buscador semántico es especializado para términos sobre biodiversidad amazónica, entonces la investigación realizada es una forma específica para implementar una web semántica o de tercera generación. Asimismo, se ha considerado que nuestro buscador semántico presenta como indicadores de evaluación más importantes a la interfaz rápida y a su tiempo de respuesta con un 25.9% y 25.5% de aceptación respectivamente, lo que nos indica que la rapidez del buscador semántico está garantizada según la evaluación realizada, sin embargo Navarro (2012) nos da a conocer que existen quince (15) características que son: formato de anotación, entorno colaborativo, diseño centrado en el usuario, formato de los documentos, consistencia de anotaciones, almacenamiento, perfil de usuario, entorno, fecha de la última versión, tipos de anotación, estructura de las anotaciones, vocabulario de anotación, papel de las ontologías, análisis de la automatización y aprendizaje; los cuales son requisitos fundamentales que deben cumplir los sistemas semánticos para ser considerados como tal, pero a la vez nos indica que a la fecha no existe ninguna herramienta que logre reunir todas esas características; por esta razón al menos nos propusimos a tener en cuenta las características más importantes para la implementación y uso del buscador, logrando así contrastar las características de una interfaz rápida con el diseño centrado en el usuario y perfil de usuario y con respecto a la rapidez en tiempo de respuesta con las características de almacenamiento, papel de las ontologías, análisis de la automatización y aprendizaje.
- Entre los aspectos relevantes descritos en la problemática destaca la denominada “infoxicación”, entendida como la sobrecarga de información y las búsquedas limitadas o ineficientes alrededor de la denominada web superficial (viendo a la web superficial

como el universo que puede ser indexado por los buscadores que según Piscitelli 500 veces inferior a la web profunda). Generando que los resultados que devuelven los buscadores no son precisamente los que buscan los usuarios y a la vez demandan consumo de tiempo. En este contexto y de acuerdo al Gráfico N° 4, es importante anotar que de las cuatro variables estudiadas Cantidad de Registros y Utilidad de Información son en las que el Buscador Semántico posee mayor superioridad en los promedios que el Buscador Simple. El Buscador Semántico al estar basado en significados (Corneliu, 2005), en la práctica emplea N criterios buscando un mismo concepto (relacionado en la ontología), por lo que es probable que se obtenga mayor cantidad de resultados. A la vez el hecho de que la variable Utilidad de Información sea la segunda también favorece el acceso a información de calidad, mejorándose las cifras de la satisfacción de lo que buscan los usuarios.

- En la actualidad la atención por la sensibilidad y validez de conclusión estadística del diseño de la investigación ha aumentado, especialmente en el tratamiento que reciben en las ediciones actuales de los manuales de diseños experimentales aunque quizás en el ámbito aplicado (donde la estimación del tamaño del efecto cobra su mayor importancia) en esta investigación se determinó el tamaño del efecto de la eficiencia en los resultados de las búsquedas de información sobre biodiversidad amazónica al utilizar el buscador semántico en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014 se empleó la ecuación (4.1) propuesta por Flores N. et al (2014) con lo que se logró el objetivo específico de la investigación: Determinar el tamaño del efecto de la eficiencia en los resultados de las búsquedas de información sobre biodiversidad amazónica al utilizar el buscador semántico en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014

Al determinar el tamaño del efecto que es de 0.7321 (73.21%), significa que la eficiencia en los resultados de las búsquedas de información sobre biodiversidad amazónica es grande, al utilizar el buscador simple y semántico respectivamente en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014. Sin embargo Flores N. et al (2014), menciona que conviene tener claro desde el principio que el valor de la estimación del tamaño del efecto debe ser interpretado en el contexto de un estudio y área concreta de investigación ya que un pequeño tamaño del efecto puede ser de gran importancia práctica en un contexto concreto por ejemplo de intervención clínica.

Al realizar el análisis inferencial a través de la prueba estadística inferencial de t student (t) se encontró que $T_C = -7.524$, $T_\alpha = -1.895$, $gl = 0.05$, $\alpha = 0.05$, es decir $T_\alpha > T_C$ rechazando la hipótesis nula (H_0) y la hipótesis alterna es aceptada (H_a), por lo tanto se prueba la validez de la hipótesis con un nivel de error de 5% siendo: La implementación de un buscador semántico incrementa significativamente la eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014.

VI. Conclusiones y recomendaciones

6.1. Conclusiones

1. Se implementó un buscador semántico con la cual se incrementó la eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014.
2. La implementación y uso de un buscador semántico sobre biodiversidad amazónica es muy aceptable (52%) a través de sus indicadores: interfaz rápida, secuencia a seguir, tiempo de respuesta y listado de registro en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014.
3. La eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica es muy aceptable (48%) en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014.
4. El tamaño del efecto de la eficiencia en los resultados de las búsquedas de información sobre biodiversidad amazónica es grande (73.21%) en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana en el año 2014

6.2. Recomendaciones

1. Continuar actualizando la ontología para obtener la mayor cantidad posible de especies sobre la biodiversidad amazónica, ya que la presente investigación solamente cuenta con 370 especies, que fueron ingresadas como muestra para su desarrollo.
2. Esta investigación sirva como guía para que se puedan crear más ontologías dedicadas a otros temas, de manera que vayamos implementando la web 3 y ser parte de ella, sobre todo en temas de nuestra región.
3. Seguir evaluando ésta investigación con las actualizaciones que se indican en la primera recomendación con la finalidad de garantizar su eficiencia, y no solo con los estudiantes de la facultad de biología sino también con los de agronomía y forestal.

Referencias bibliográficas

Beltré, H. (2008). Aplicación de la Usabilidad al Proceso de desarrollo de Páginas Web. Tesis de Master en Tecnologías de la Información. Universidad Politécnica de Madrid. Facultad de Informática. p. 22. Extraído del Word Wide Web: http://oa.upm.es/1176/1/HAYSER_JACQUELIN_BELTRE_FERRERAS.pdf

Berners-Lee, T. (2005), Weaving a Semantic Web, Extraído del Word Wide Web: <http://www.digitaldivide.net/articles/view.php?ArticleID=20>

Berners-Lee, T., Hendler, J. y Lassila, O. (2001). The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, Mayo, pp. 28-37. Extraído del World Wide Web: <http://www.sciam.com/article.cfm?id=the-semantic-web>.

Castells, P. (2003). La web semántica, Escuela Politécnica Superior Universidad Autónoma de Madrid, España. p. 5. Extraído del World Wide Web: <http://arantxa.ii.uam.es/~castells/publications/castells-uclm03.pdf>

Contreras, J. y Martínez, J. (2007). Tutorial de Ontologías, Universidad Complutense de Madrid, España. pp. 7 – 8 Extraído del World Wide Web: http://www.sedic.es/gt_normalizacion_tutorial_ontologias.pdf

Corneliu, B. S. (2005). Semantic Web: Fundamente si aplicatii.

Criado, F., L. (2009). Procedimiento semi – automático para transformar la Web en Web Semántica. Tesis de Doctorado. Universidad Nacional de Educación a Distancia (UNED), Escuela Técnica Superior de Ingeniería Informática, Departamento de Inteligencia Artificial.

Cueva, S. (2008). Web Semantica: Principios y estándares.

Efigenia, A., Sandoval, C. y Caldas, F. Uso de ontologías y web semántica para apoyar la gestión del conocimiento. Grupo de Investigación Gestión Empresarial e Innovación

Tecnológica GEIT, Universidad Distrital Francisco José de Caldas. Universidad Militar Nueva Granada, Bogotá COLOMBIA.

Escuela de sistemas (2011). Una ontología para la representación de conceptos de diseño de software. Universidad Nacional de Colombia.

Farquhar, A. (1997). Ontolingua Tutorial. University of Stanford. Extraído del Word Wide Web: <http://www.ksl.stanford.edu/software/ontolingua/tutorial.pdf>

Flores N. (2014). Análisis de datos para diseños de dos grupos. Muestras independientes. Web: <http://www.psicocode.com/resumenes/tema3disenos.pdf>

García, F. (2007). Paper: “Web Semántica y Ontologías”, Departamento de Informática y Automática – Facultad de Ciencias Universidad de Salamanca. Extraído del Word Wide Web: <http://misiondigital.net/cms/wp-content/uploads/2014/04/10-WSemantica.pdf>

Gruber, T. (2003). Extraído del Word Wide Web:
<http://www.infor.uva.es/~sblanco/Tesis/Ontologías.pdf>

Grüninger, M. y Fox, M. S. (1995b). The Logic of Enterprise Modelling. En J. Brown, D. O’Sullivan (Eds.), *Reengineering the Enterprise* (pp. 83-98). Chapman and Hall.

Hendler, J., Berners-Lee, T. y Miller, E. (2002). Integrating applications on the semantic web. *Journal of the Institute of Electrical Engineers of Japan*, 122(10) p.p 676-680. Extraído del World Wide Web: <http://www.w3.org/2002/07/swint>

Lopez, C., García F. (2005). Los contenidos educativos y la Web semántica. Facultad de Informática y Automática Universidad De Salamanca, España. p. 3.

Lozano, T., A. (2002). Métrica de Idoneidad de Ontologías. Tesis de Doctorado. Universidad de Extremadura. Departamento de Informática.

Mizoguchi, R., Ikeda, M. y Sinitsa, K. (1997). Roles of Shared Ontology in AI-ED Research. En B. de Boulay, R. Mizoguchi (Eds.), *Artificial Intelligence in Education AI-ED* 97 (pp. 537-544). IOS Press.

Muñoz, P., J. (2009). Teoría de Modelado del E-Learning y Aplicación a un Sistema de Pistas adaptativo en Tutoría inteligente utilizando técnicas de Web Semántica. Tesis de Doctorado. Universidad Carlos III de Madrid. Departamento de Ingeniería Telemática.

Navarro, G., J. (2012). FLERSA: Un sistema semántico de Gestión de contenido Web (S-CMS), Tesis de Doctorado. Universidad de Granada. Departamento de Lenguajes y Sistemas Informáticos.

Noriega, G., M. (2009). Modelado de Sistemas CSCW siguiendo un enfoque de ingeniería dirigida por ontologías, Tesis de Doctorado. Universidad de Granada. Departamento de Lenguajes y Sistemas Informáticos.

Noy, N., F. & McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. Stanford University, Stanford, Extraído del World Wide Web: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf

Ordóñez, (2007). Aportes de las Ontologías en la recuperación de la web (2007). Universidad Mariana, Facultad de ingeniería. COLOMBIA.

Palacios, M., V. (2010). Sistema de Recuperación conceptual mediante niveles semánticos en la representación de esquemas de metadatos, tesis de doctorado. Universidad Carlos III de Madrid. Departamento de Biblioteconomía y Documentación

Peinado, Federico. Un armazón para el desarrollo de aplicaciones de narración automática para el desarrollo de aplicaciones de narración automática basado en componentes ontológicos. Tesis Doctoral. Universidad Complutense de Madrid. España.

Portal Acción TI (2013).

<http://www.accionti.com/blog/2013/01/semantica-en-los-buscadores/>

Portal del Sistema de Información de la Diversidad Biológica y Ambiental de la Amazonia Peruana – SIAMAZONIA

www.siamazonia.org.pe

Portal de la Fundación EROSKI

<http://www.consumer.es/web/es/tecnologia/internet/2004/12/22/114125.php>.

Portal de Universidad Nacional del Nordeste de Argentina

<http://www.unne.edu.ar/>

Reyes G., S. (2007). Propuesta para el modelado del conocimiento empresarial, Tesis Doctoral. Universitat Jaume I de Castello, España.

Rivera, E., Maturana. SISDEON: Un sistema de información sobre el desplazamiento forzado en Colombia mediante el uso de ontologías. Universidad Nacional de Colombia Sede Medellín, Colombia.

Romero L., R (2007). Especificación OWL de una ontología para teleeducación en la Web Semántica, tesis doctoral. Universidad Politécnica de Valencia, Departamento de Comunicaciones.

Torres R., N. (2008). Sistemas de análisis automático de fotografías. Modelo Conceptual según los estándares de la Web Semántica. Tesis Doctoral. Universidad Carlos III de Madrid.

Sitio Web Acción TI, 2003

<http://www.accionti.com/paginas-web.html>

Sitio Web Buscadores Semánticos de Guiomar Pérez López (2010)

<https://sites.google.com/site/buscadoressemanticos/-que-es-un-buscador-semantico>

Sitio Web Informatica 1101

<http://informatica1101.bligoo.com/>

Sowa, J. F. (2000). Knowledge Representation: Logical, Philosophical, and Computational Foundations. Pacific Grove, CA: Brooks Cole Publishing Co. Extraído del World Wide Web: <http://www.aclweb.org/anthology/J01-2006>

Studer S., Benjamins R. y Fensel D. (1998). "Knowledge Engineering: Principles and Methods", *Data and Knowledge Engineering*, p. 25.

Uschold, M. y Grüninger, M. (1996) Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review*, 11(2), p.p. 93-155.

Yusef Hassan & Francisco J. Martín Fernández & Ghzala Iazza. (2004). Diseño Web Centrado en el Usuario: Usabilidad y Arquitectura de la Información. Universidad de Granada. Extraído del World Wide Web: http://www.upf.edu/hipertextnet/numero-2/disenio_web.html.

World Wide Web Consortium - W3C

<http://www.w3c.es/>

Anexos

Anexo N° 01: Matriz de consistencia

Anexo N° 02: Cuestionarios (pre test y post test)

ANEXO Nº 01: Matriz de Consistencia

TITULO: EFICIENCIA EN BÚSQUEDAS DE INFORMACIÓN SOBRE BIODIVERSIDAD AMAZONICA CON EL USO DE UN BUSCADOR SEMÁNTICO- UNIVERSIDAD NACIONAL DE LA AMAZONIA PERUANA, 2014

PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES	INDICADORES	METODOLOGÍA
<p>Problema General</p> <p>¿Cómo el uso de un buscador semántico incrementa la eficiencia de las búsquedas de información sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonia Peruana en el año 2014?</p> <p>Problemas Específicos</p> <p>¿Cómo es la evaluación del uso de un buscador semántico sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonia Peruana en el año 2014?.</p> <p>¿Cómo es la eficiencia de los resultados de las búsquedas sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonia Peruana en el año 2014?.</p> <p>¿Cuál es el tamaño del efecto de la eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica al utilizar el buscador simple y semántico respectivamente en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonia Peruana en el año 2014?.</p>	<p>Objetivo General</p> <p>Usar un buscador semántico para incrementar la eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonia Peruana en el año 2014.</p> <p>Objetivos Específicos</p> <ul style="list-style-type: none"> • Evaluar el uso de un buscador semántico sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonia Peruana en el año 2014. • Evaluar la eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonia Peruana en el año 2014. • Determinar el tamaño del efecto de la eficiencia en los resultados de las búsquedas de información sobre biodiversidad amazónica al utilizar el buscador semántico en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonia Peruana en el año 2014. 	<p>Hipótesis Alterna</p> <p>El uso de un buscador semántico incrementara significativamente la eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonia Peruana en el año 2014.</p> <p>Hipótesis Nula</p> <p>El uso de un buscador semántico no incrementara significativamente la eficiencia en los resultados de las búsquedas sobre biodiversidad amazónica en los estudiantes de Ciencias Biológicas de la Universidad Nacional de la Amazonia Peruana en el año 2014.</p>	<p>VARIABLE INDEPENDIENTE (X)</p> <p>Buscador semántico.</p> <p>VARIABLE DEPENDIENTE (Y)</p> <p>Eficiencia de las búsquedas de información sobre biodiversidad</p>	<p>X1. Tiempo de apertura de interfaz.</p> <p>X2. Proceso a seguir para la búsqueda.</p> <p>X3 Tiempo de respuesta a la búsqueda.</p> <p>X4. Listado de resultados de búsquedas.</p> <p>Y1. Tiempo total para realizar una búsqueda</p> <p>Y2. Cantidad de pasos realizados de inicio a fin.</p> <p>Y3. Cantidad de registros encontrados.</p> <p>Y4. Utilidad de los registros encontrados (información útil).</p>	<p>Tipo de Investigación</p> <p>La investigación es de tipo descriptivo.</p> <p>Diseño de la Investigación</p> <p>La investigación pertenece al diseño cuasiexperimental.</p> <p>El esquema es: G₁ : O₁ X O₂</p> <p>Dónde: G₁ : Grupo experimental X : Tratamiento con el prototipo O₁ : Test antes del experimento O₂ : Test después del experimento</p> <p>Población y Muestra:</p> <p>-Población: 250 estudiantes de Ciencias Biológicas</p> <p>-Muestra: 50 estudiantes de Ciencias Biológicas</p> <p>- Técnicas e instrumentos de recolección de datos</p> <p>Técnicas: Encuestas</p> <p>Instrumentos: Cuestionario</p>



UNIVERSIDAD NACIONAL DE LA AMAZONÍA PERUANA

Escuela de Post Grado

MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN GERENCIA DE TIC Y GESTIÓN DE SOFTWARE

ANEXO Nº 02: Cuestionario (Post Test)

TÍTULO

EFICIENCIA EN LAS BÚSQUEDAS DE INFORMACIÓN SOBRE BIODIVERSIDAD AMAZONICA CON EL USO DE UN BUSCADOR SEMÁNTICO EN LOS ESTUDIANTES DE CIENCIAS BIOLÓGICAS DE LA UNIVERSIDAD NACIONAL DE LA AMAZONÍA PERUANA EN EL AÑO 2014

(Para estudiantes de la Facultad de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana)

I. PRESENTACIÓN.

¡BUENOS DIAS!

- El presente cuestionario tiene como propósito obtener información sobre resultados de búsqueda en temas de biodiversidad con el uso de un Buscador Semántico.
- Los sujetos seleccionados para el estudio son los estudiantes matriculados en el II semestre académico 2014, de la Facultad de Ciencias Biológicas
- Este estudio servirá para elaborar la tesis conducente a la obtención del grado académico de “Magister en Ingeniería de sistemas”

II. DATOS GENERALES DEL ESTUDIANTE

1. FACULTAD:.....
2. NOMBRE Y APELLIDO:.....
3. CÓDIGO:.....
3. NIVEL :.....
4. CICLO :.....
5. EDAD:.....
6. SEXO:.....
7. FECHA:.....

III. INSTRUCCIONES.

- Responda a las preguntas que se encuentran en el cuestionario que no llevará mucho tiempo.
- La información que nos proporcione será manejada con la más estricta confidencialidad.
- Responda a todas las preguntas con la mayor sinceridad que el caso requiere.
- No deje enunciados sin responder.
- Lea en orden cada uno de los enunciados y marque con un aspa (X) en aquella columna que según usted , se acerca más a su realidad como estudiante

IV. CONTENIDO

A usted se le ha presentado la oportunidad de utilizar el buscador denominado “buscador semántico”. Utilice los mismos 3 criterios de búsqueda sobre biodiversidad que utilizó en el test anterior, revise los resultados y responda las siguientes preguntas:

Leyenda:

- Muy apropiado (5)
- Ligeramente apropiado (4)
- Apropiado (3)
- Ligeramente inapropiado (2)
- Muy inapropiado (1)

Preguntas	Puntuación	Sistema de Puntuaciones				
		(5)	(4)	(3)	(2)	(1)
1. Sobre el tiempo de apertura de interfaz: ¿Considera usted que el tiempo para visualizar la interfaz del buscador es apropiado para los usuarios?						
2. Sobre el proceso a seguir para realizar una búsqueda: ¿Considera usted que la secuencia seguida en el buscador para realizar una búsqueda es apropiado para los usuarios?						
3. Sobre el tiempo de respuesta a la búsqueda: ¿Considera usted que el tiempo de respuesta del buscador para encontrar y listar los resultados es apropiado para los usuarios?						
4. Sobre el listado de registros de búsqueda: ¿Considera usted que el listado de registros (visibilidad y entendimiento del resultado) del buscador es apropiado para los usuarios?						
5. Sobre el tiempo que utilizó para realizar en una búsqueda: ¿Considera usted que el tiempo total que necesitó para realizar su búsqueda es apropiado?						
6. Sobre la búsqueda realizada: ¿Considera usted que la cantidad de pasos realizados en el buscador desde la apertura de interfaz hasta encontrar los resultados es apropiado?						
7. Sobre la cantidad de registros encontrados: ¿Considera usted que la cantidad de registros encontrados por el buscador es apropiado?						
8. Sobre la utilidad de los registros encontrados: ¿Considera usted que los contenidos (información) de los registros encontrados es apropiado?						



UNIVERSIDAD NACIONAL DE LA AMAZONÍA PERUANA
Escuela de Post Grado

**MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN GERENCIA DE
TIC Y GESTIÓN DE SOFTWARE**

Anexo N° 02: Cuestionario (Pre Test)

TÍTULO

EFICIENCIA EN LAS BÚSQUEDAS DE INFORMACIÓN SOBRE BIODIVERSIDAD AMAZONICA CON EL USO DE UN BUSCADOR SEMÁNTICO EN LOS ESTUDIANTES DE CIENCIAS BIOLÓGICAS DE LA UNIVERSIDAD NACIONAL DE LA AMAZONÍA PERUANA EN EL AÑO 2014

(Para estudiantes de la Facultad de Ciencias Biológicas de la Universidad Nacional de la Amazonía Peruana)

I. PRESENTACIÓN.

¡BUENOS DIAS!

- El presente cuestionario tiene como propósito obtener información sobre resultados de búsqueda en temas de biodiversidad con el uso de un Buscador Semántico.
- Los sujetos seleccionados para el estudio son los estudiantes matriculados en el II semestre académico 2014, de la Facultad de Ciencias Biológicas
- Este estudio servirá para elaborar la tesis conducente a la obtención del grado académico de “Magister en Ingeniería de sistemas”

II. DATOS GENERALES DEL ESTUDIANTE

1. FACULTAD:.....
2. NOMBRE Y APELLIDO:.....
3. CÓDIGO:.....
3. NIVEL :.....
4. CICLO :.....
5. EDAD:.....
6. SEXO:.....
7. FECHA:.....

III. INSTRUCCIONES.

- Responda a las preguntas que se encuentran en el cuestionario que no llevará mucho tiempo.
- La información que nos proporcione será manejada con la más estricta confidencialidad.
- Responda a todas las preguntas con la mayor sinceridad que el caso requiere.
- No deje enunciados sin responder.
- Lea en orden cada uno de los enunciados y marque con un aspa (X) en aquella columna que según usted , se acerca más a su realidad como estudiante

IV. CONTENIDO

A usted se le ha presentado la oportunidad de utilizar el buscador denominado “buscador semántico”. Utilice los mismos 3 criterios de búsqueda sobre biodiversidad que utilizó en el test anterior, revise los resultados y responda las siguientes preguntas:

Leyenda:

- Muy apropiado (5)
- Ligeramente apropiado (4)
- Apropiado (3)
- Ligeramente inapropiado (2)
- Muy inapropiado (1)

Preguntas	Puntuación	Sistema de Puntuaciones				
		(5)	(4)	(3)	(2)	(1)
1. Sobre el tiempo de apertura de interfaz: ¿Considera usted que el tiempo para visualizar la interfaz del buscador es apropiado para los usuarios?						
2. Sobre el proceso a seguir para realizar una búsqueda: ¿Considera usted que la secuencia seguida en el buscador para realizar una búsqueda es apropiado para los usuarios?						
3. Sobre el tiempo de respuesta a la búsqueda: ¿Considera usted que el tiempo de respuesta del buscador para encontrar y listar los resultados es apropiado para los usuarios?						
4. Sobre el listado de registros de búsqueda: ¿Considera usted que el listado de registros (visibilidad y entendimiento del resultado) del buscador es apropiado para los usuarios?						
5. Sobre el tiempo que utilizó para realizar en una búsqueda: ¿Considera usted que el tiempo total que necesitó para realizar su búsqueda es apropiado?						
6. Sobre la búsqueda realizada: ¿Considera usted que la cantidad de pasos realizados en el buscador desde la apertura de interfaz hasta encontrar los resultados es apropiado?						
7. Sobre la cantidad de registros encontrados: ¿Considera usted que la cantidad de registros encontrados por el buscador es apropiado?						
8. Sobre la utilidad de los registros encontrados: ¿Considera usted que los contenidos (información) de los registros encontrados es apropiado?						