

**UNIVERSIDAD PERUANA DE CIENCIAS E INFORMÁTICA**  
**FACULTAD DE CIENCIAS E INGENIERÍA**  
**CARRERA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA**



**TESIS:**

**“Implementación de una Aplicación Web Scraping y la Mejora de la Publicación de Noticias Tecnológicas en el Portal Web Ojotecnológico”**

**AUTOR:**

**Bach. Chuquijajas Chuquilín, Jorge Luis**

**PARA OPTAR EL TÍTULO PROFESIONAL DE**  
**Ingeniero De Sistemas E Informática**

**ASESOR:**

**Mg. Hermoza Ochante, Rubén Edgar**

Código ORCID: 0000-0003-4769-0101

DNI N°: 42037740

**LIMA – PERÚ**  
**2023**



**INFORME DE SIMILITUD N°006-2022- REHO-A**

**A** : **MG. CESAR QUISPE AYQUIPA**  
Decano (e) de la Facultad de Ciencias e Ingeniería

**DE** : **MG. HERMOZA OCHOA RUBÉN EDGAR**  
Docente Operador del Programa Turnitin

**ASUNTO** : Informe de evaluación de Similitud de Tesis.  
**BACHILLER JORGE LUIS CHUQUIAJAS CHUQUILÍN**

**FECHA** : Lima, 18 de noviembre del 2022

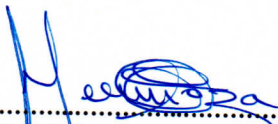
---

Tengo el agrado de dirigirme a usted con la finalidad de informar lo siguiente:

1. Mediante el uso del programa informático **Turnitin** (con las configuraciones de excluir citas, excluir bibliografía y excluir oraciones con cadenas menores a 15 palabras) se ha analizado la tesis titulada: "**IMPLEMENTACIÓN DE UNA APLICACIÓN WEB SCRAPING Y LA MEJORA DE LA PUBLICACIÓN DE NOTICIAS TECNOLÓGICAS EN EL PORTAL WEB OJOTECNOLOGICO**", presentado por Bachiller **JORGE LUIS CHUQUIAJAS CHUQUILÍN**.
2. Los resultados de la evaluación concluyen que el bachiller en mención tiene un **ÍNDICE DE SIMILITUD DE 29%** (cumpliendo con el artículo 35 del Reglamento de Grado de Bachiller y Título Profesional UPCI aprobado con Resolución N° 373-2019-UPCI-R de fecha 22/08/2019).
3. Al término análisis, el Bachiller en mención **PUEDEN CONTINUAR** su trámite ante la facultad, por lo que el resultado del análisis se adjunta para los efectos consiguientes

Es cuanto hago de conocimiento para los fines que se sirva determinar.

Atentamente,

  
.....  
**MG. HERMOZA OCHOA RUBÉN EDGAR**  
Universidad Peruana de Ciencias e Informática  
Docente Operador del Programa Turnitin

## **DEDICATORIA**

A mi mamá en que se fue al cielo en el desarrollo de mi tesis, y que fue el motor que me impulso a hacer profesional.

A mi papá y mi hermano que siempre cuento con su apoyo incondicional.

A mi tía Marina que me albergo en su casa desde el día tome la decisión de buscar un futuro mejor y por su apoyo incondicional.

A mis amigos Anthony y Juan con quienes desarrollamos ojotecnologico, con quien me une gran amistad.

## **AGRADECIMIENTOS**

A Dios por todo lo que me brinda cada día.

A mi mamá que es mi ángel y sé que me guiara todos los días, y por grandiosos momentos que pasamos en familia.

A mi papá y hermano que siempre estarán conmigo para todo lo que me proponga en mi vida.

A mi tía Marina que es mi segunda madre al brindarme todo el apoyo para lograr mis objetivos.

A mi asesor el Mg. HERMOZA OCHANTE RUBÉN EDGAR, por la paciencia y todos los conocimientos brindados.

## PRESENTACIÓN

La investigación analiza las teorías primordiales, describe y revela la percepción sobre la web scraping y la automatización de los procesos para la obtención de noticias tecnológicas de diferentes fuentes de comunicación. El cual nos permite sentar las bases de solución y adaptación de mejoras en los procesos de alguna realiza similar a la que se a planteando en la investigación.

El estudio que se realizo es importante para los administradores del portal de noticias tecnológicas ojotecnologico, el cual les hará entender la importancia de la automatización de procesos y que ha permitido significativamente la reducción de tiempo y costo.

Cabe mencionar que el portal web ojotecnológico es parte de un emprendimiento de jóvenes que quiere lograr formalizar el proyecto convirtiendo lo desarrollado en un servicio para que empresas lo puedan utilizar.

Los estudios similares que se pudo obtener a la realizar la investigación son muy pocos, ya que la mayoría están enfocados a utilizar la web scraping para el análisis de comparación de productos, análisis de competencia, SEO, el cual permiten tomar decisiones, pero en nuestra investigación demostramos que se puede automatizar procesos para que el usuario final sea el beneficiado.

Para obtener los principales objetivos de la investigación se lleva a cabo en 6 capítulos:

En capítulo I, se presenta la Realidad problemática, Planteamiento del problema, Hipótesis de la investigación, Objetivos de la investigación, Variables, dimensiones e

indicadores, Justificación del estudio, detallando los Trabajos Previos, Teorías relacionadas al tema, Definición de términos básicos y Limitaciones de la Investigación.

En capítulo II, se expone la Metodología, Tipo y Diseño de investigación, Población, Muestra y muestreo, Técnicas e instrumentos para la recolección de datos, Valides y Confiabilidad de instrumentos, Método de análisis de datos y Aspectos Éticos.

En capítulo III, se desarrolla la presentación de resultados, Resultados descriptivos, Prueba de normalidad y la Contrastación de las Hipótesis de la investigación.

En capítulo IV, se plantea la discusión de los resultados obtenidos de la hipótesis general y de las hipótesis específicas.

En capítulo V, se plantean las conclusiones de la investigación a partir de los resultados favorables de las hipótesis planteadas para demostrar que la web scraping si mejoró el proceso de publicación de noticias

En capítulo VI, se plantean las recomendaciones y oportunidad de mejora que pueden ser implementadas progresivamente en el portal web ojotecnologico.

En la parte final se presenta las referencias bibliográficas y los anexos que complementan y ampara el estudio de la investigación.

## ÍNDICE

<b>DEDICATORIA .....</b>	<b>ii</b>
<b>AGRADECIMIENTOS.....</b>	<b>iii</b>
<b>PRESENTACIÓN .....</b>	<b>iv</b>
<b>ÍNDICE.....</b>	<b>vi</b>
<b>ÍNDICE DE TABLAS .....</b>	<b>viii</b>
<b>ÍNDICE DE FIGURAS .....</b>	<b>xi</b>
<b>RESUMEN .....</b>	<b>xiii</b>
<b>ABSTRACT .....</b>	<b>xiv</b>
<b>I. Introducción .....</b>	<b>1</b>
<b>1.1 Realidad problemática .....</b>	<b>1</b>
<b>1.2 Planteamiento del problema .....</b>	<b>7</b>
1.2.1 Problema general .....	7
1.2.2 Problemas Específicos .....	7
1.2.3 Delimitación del problema de investigación Espacial.....	7
1.2.4 Delimitación del problema de investigación Temporal.....	8
<b>1.3 Hipótesis de la investigación .....</b>	<b>8</b>
1.3.1 Hipótesis General .....	8
1.3.2 Hipótesis Específicas .....	8
<b>1.4 Objetivos de la investigación .....</b>	<b>8</b>
1.4.1 Objetivo General.....	8
1.4.2 Objetivos Específicos .....	9
<b>1.5 Variables, dimensiones e indicadores .....</b>	<b>9</b>
<b>1.6 Justificación del estudio. ....</b>	<b>10</b>
1.6.1 Justificación Teórica.....	10
1.6.2 Justificación Práctica .....	10
1.6.3 Justificación Metodológica.....	10
1.6.4 Justificación Social .....	11
1.6.5 Justificación Legal .....	11
1.6.6 Justificación Económica .....	11
1.6.7 Importancia.....	11
<b>1.7 Trabajos Previos. ....</b>	<b>15</b>
<b>1.8 Teorías relacionadas al tema .....</b>	<b>24</b>
<b>1.9 Definición de términos básicos. ....</b>	<b>45</b>
<b>II. MÉTODO.....</b>	<b>50</b>
<b>2.1 Tipo y Diseño de investigación .....</b>	<b>50</b>
2.1.1 Diseño de la investigación.....	50
2.1.2 Tipo de investigación.....	50
2.1.3 Nivel de la investigación .....	50
2.1.4 Enfoque de la investigación.....	51
<b>2.2 Población, Muestra y muestreo .....</b>	<b>51</b>
2.2.1 Población .....	51

2.2.2	Muestra .....	52
2.2.3	Muestreo .....	53
<b>2.3</b>	<b>Técnicas e instrumentos para la recolección de datos.....</b>	<b>53</b>
2.3.1	Técnicas .....	53
2.3.2	Instrumentos .....	54
<b>2.4</b>	<b>Valides y Confiabilidad de instrumentos.....</b>	<b>54</b>
<b>2.5</b>	<b>Método de análisis de datos .....</b>	<b>57</b>
<b>2.6</b>	<b>Aspectos Éticos.....</b>	<b>58</b>
<b>III.</b>	<b>RESULTADOS .....</b>	<b>59</b>
<b>3.1</b>	<b>Resultados descriptivos .....</b>	<b>59</b>
3.1.1	Descripción del proceso manual de publicación .....	59
3.1.2	Descripción de la implementación web scraping .....	64
3.1.3	Análisis e interpretación de resultados de web scraping .....	75
3.1.4	Análisis e interpretación de resultados de publicación de noticias .....	89
<b>3.2</b>	<b>Prueba de Normalidad .....</b>	<b>103</b>
<b>3.3</b>	<b>Contrastación de las Hipótesis.....</b>	<b>104</b>
3.3.1	Contrastación de la Hipótesis General.....	104
3.3.2	Contrastación de Hipótesis Específicas .....	107
<b>IV.</b>	<b>DISCUSIÓN.....</b>	<b>113</b>
<b>V.</b>	<b>CONCLUSIONES .....</b>	<b>114</b>
<b>VI.</b>	<b>RECOMENDACIONES.....</b>	<b>115</b>
	<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>116</b>
	<b>ANEXOS.....</b>	<b>119</b>
	<b>Anexo 01: Matriz de Consistencia.....</b>	<b>119</b>
	<b>Anexo 02: Instrumentos de recolección de datos .....</b>	<b>120</b>
	<b>Anexo 03: Base de datos.....</b>	<b>121</b>
	<b>Anexo 04: Evidencia de similitud digital .....</b>	<b>127</b>
	<b>Anexo 05: Autorización de publicación en repositorio .....</b>	<b>129</b>



## ÍNDICE DE TABLAS

Tabla 1 Fuentes de Comunicación.....	2
Tabla 2 Variables, dimensiones e indicadores.....	9
Tabla 3 Sistemas de extracción de información web disponibles comercialmente.....	13
Tabla 4: Método de web scraping, aplicación a utilizar y facilidad. ....	31
Tabla 5 Servicios de Hosting.....	33
Tabla 6 Tipos de Hosting.....	34
Tabla 7 Técnicas de recolección de datos. ....	53
Tabla 8 Instrumento de recolección de datos. ....	54
Tabla 9 Fiabilidad Alfa de Cronbach .....	55
Tabla 10 Fiabilidad Alfa de Cronbach total de los elementos.....	55
Tabla 11 Matriz análisis de Datos. ....	57
Tabla 12 Situación actual de número de noticias por día.....	63
Tabla 13 Cantidad de noticias diarias luego de aplicar web Scraping .....	68
Tabla 14 Cantidad de noticias diarias luego de aplicar web Scraping .....	71
Tabla 15 Resultado Descriptivo sobre la pregunta uno 1 (Pre Test).....	75
Tabla 16 Resultado Descriptivo sobre la pregunta uno 1 (Post Test) .....	76
Tabla 17 Resultado Descriptivo sobre la pregunta 2 (Pre Test) .....	77
Tabla 18 Resultado Descriptivo sobre la pregunta 2 (Post Test) .....	77
Tabla 19 Resultado Descriptivo sobre la pregunta 3 (Pre Test) .....	78
Tabla 20 Resultado Descriptivo sobre la pregunta 3 (Post Test) .....	78
Tabla 21 Resultado Descriptivos sobre la pregunta 4 (Pre Test) .....	79
Tabla 22 Resultados Descriptivo sobre la pregunta 4 (Post Test).....	80
Tabla 23 Resultado Descriptivo sobre la pregunta 5 (Pre Test).....	81
Tabla 24 Resultado Descriptivo sobre la pregunta 5 (Post Test) .....	81
Tabla 25 Resultados Descriptivos sobre la pregunta 6 (Pre Test).....	82
Tabla 26 Resultado Descriptivo sobre la pregunta 6 (Post Test) .....	82
Tabla 27 Resultado Descriptivo sobre la pregunta 7 (Pre Test).....	83
Tabla 28 Resultado Descriptivo sobre la pregunta 7 (Post Test) .....	83
Tabla 29 Resultado Descriptivo sobre la pregunta 8 (Pre Test).....	84
Tabla 30 Resultado Descriptivo sobre la pregunta 8 (Post Test) .....	84
Tabla 31 Resultado Descriptivo sobre la pregunta 9 (Pre Test).....	85

Tabla 32 Resultado Descriptivo sobre la pregunta 9 (Post Test) .....	85
Tabla 33 Resultado Descriptivo sobre la pregunta 10 (Pre Test) .....	86
Tabla 34 Resultado Descriptivo sobre la pregunta 10 (Post Test) .....	86
Tabla 35 Resultado Descriptivo sobre la pregunta 11 (Pre Test) .....	87
Tabla 36 Resultados Descriptivos sobre la pregunta 11 (Post Test) .....	87
Tabla 37 Resultado Descriptivo sobre la pregunta 12 (Pre Test) .....	88
Tabla 38 Resultado Descriptivo sobre la pregunta 12 (Post Test) .....	88
Tabla 39 Resultado Descriptivo sobre la pregunta 13 (Pre Test) .....	89
Tabla 40 Resultado Descriptivo sobre la pregunta 13 (Post Test) .....	89
Tabla 41 Resultado Descriptivo sobre la pregunta 14 (Pre Test) .....	90
Tabla 42 Resultado Descriptivo sobre la pregunta 14 (Post Test) .....	91
Tabla 43 Resultado Descriptivo sobre la pregunta 15 (Pre Test) .....	92
Tabla 44 Resultado Descriptivo sobre la pregunta 15 (Post Test) .....	92
Tabla 45 Resultado Descriptivo sobre la pregunta 16 (Pre Test) .....	93
Tabla 46 Resultado Descriptivo sobre la pregunta 16 (Post Test) .....	93
Tabla 47 Resultado Descriptivo sobre la pregunta 17 (Pre Test) .....	94
Tabla 48 Resultados Descriptivos sobre la pregunta 17 (Post Test) .....	94
Tabla 49 Resultados Descriptivos sobre la pregunta 18 (Pre Test) .....	95
Tabla 50 Resultados Descriptivos sobre la pregunta 18 (Post Test) .....	95
Tabla 51 Resultados Descriptivos sobre la pregunta 19 (Pre Test) .....	96
Tabla 52 Resultados Descriptivos sobre la pregunta 19 (Post Test) .....	96
Tabla 53 Resultado Descriptivo sobre la pregunta 20 (Pre Test) .....	97
Tabla 54 Resultado Descriptivo sobre la pregunta 20 (Post Test) .....	97
Tabla 55 Resultado Descriptivo sobre la pregunta 21 (Pre Test) .....	98
Tabla 56 Resultado Descriptivo sobre la pregunta 21 (Post Test) .....	98
Tabla 57 Resultado Descriptivo sobre la pregunta 22 (Pre Test) .....	100
Tabla 58 Resultado Descriptivo sobre la pregunta 22 (Post Test) .....	100
Tabla 59 Resultados Descriptivos sobre la pregunta 23 (Pre Test) .....	101
Tabla 60 Resultado Descriptivo sobre la pregunta 23 (Post Test) .....	101
Tabla 61 Resultado Descriptivo sobre la pregunta 24 (Pre Test) .....	102
Tabla 62 Resultado Descriptivo sobre la pregunta 24 (Post Test) .....	102
Tabla 63 Prueba de normalidad del PRE y POST TEST de la VI y VD .....	104
Tabla 64 Resultados de la prueba t para Web Scraping .....	105
Tabla 65 Resultados de la prueba t para publicación de noticias .....	106

Tabla 66 Resultados de la prueba t para prueba de HE 1. ....	108
Tabla 67 Resultados de la prueba t para prueba de HE 2. ....	109
Tabla 68 Resultados de la prueba t para prueba de HE 3. ....	111

## ÍNDICE DE FIGURAS

Figura 1: Cantidad de noticias publicadas por día y medio de comunicación .....	4
Figura 2: ¿Cuál es el nivel de satisfacción general con ojotecnologico? .....	5
Figura 3: ¿Cree que volverá a visitarnos en los próximos 15 días? .....	6
Figura 4: Modelo general para la formulación y uso de cuestionarios.....	23
Figura 5: Representación gráfica del concepto de proceso .....	25
Figura 6: Fases de la Metodología WSDM. ....	36
Figura 7: Fases de la metodología SOHDM.....	37
Figura 8: Fases de la metodología OOHDM.....	38
Figura 9: Fases de la metodología WAE .....	39
Figura 10: Fases de la metodología IWEB .....	39
Figura 11: Comparación de requisitos en el entorno Web en las metodologías.....	41
Figura 12: Estructura del libro.....	42
Figura 13: Diagrama de flujo del proceso de publicación de noticias.....	59
Figura 14: Ciclo del proceso de publicación de noticias.....	61
Figura 15: Acceso para el administrador.....	61
Figura 16: Panel de ingreso de nuevas noticias.....	62
Figura 17: Página principal de ojotecnologico .....	63
Figura 18: Flujo del proceso de publicación de noticias con web scraping .....	65
Figura 19: Arquitectura del portal ojotecnologico. ....	66
Figura 20:Portal web ojotecnologico.....	67
Figura 21: Portal web luego de implementar el scraping. ....	70
Figura 22: Comparación de usuarios y sesiones nuevas.....	72
Figura 23: Puntaje de Aplicación web progresiva, Rendimiento, Accesibilidad .....	73
Figura 24: Informe de rendimiento de ojotecnologico. ....	74
Figura 25: Diseño Responsivo de ojotecnologico. ....	74
Figura 26: Resultados descriptivos sobre la pregunta 1 (Pre y Post) .....	76
Figura 27: Resultados descriptivos sobre la pregunta 2 (Pre y Post) .....	77
Figura 28: Resultados descriptivos sobre la pregunta 3 (Pre y Post) .....	79
Figura 29: Resultados descriptivos sobre la pregunta 4 (Pre y Post) .....	80
Figura 30: Resultados descriptivos sobre la pregunta 5 (Pre y Post) .....	81
Figura 31: Resultados descriptivos sobre la pregunta 6 (Pre y Post) .....	82
Figura 32: Resultados descriptivos sobre la pregunta 7 (Pre y Post) .....	83

Figura 33: Resultados descriptivos sobre la pregunta 8 (Pre y Post) .....	84
Figura 34: Resultados descriptivos sobre la pregunta 9 (Pre y Post) .....	85
Figura 35: Resultados descriptivos sobre la pregunta 10 (Pre y Post) .....	86
Figura 36: Resultados descriptivos sobre la pregunta 11 (Pre y Post) .....	87
Figura 37: Resultados descriptivos sobre la pregunta 12 (Pre y Post) .....	88
Figura 38: Resultados descriptivos sobre la pregunta 13 (Pre y Post) .....	90
Figura 39: Resultados descriptivos sobre la pregunta 14 (Pre y Post) .....	91
Figura 40: Resultados descriptivos sobre la pregunta 15 (Pre y Post) .....	92
Figura 41: Resultados descriptivos sobre la pregunta 16 (Pre y Post) .....	93
Figura 42: Resultados descriptivos sobre la pregunta 17 (Pre y Post) .....	94
Figura 43: Resultados descriptivos sobre la pregunta 18 (Pre y Post) .....	95
Figura 44: Resultados descriptivos sobre la pregunta 19 (Pre y Post) .....	96
Figura 45: Resultados descriptivos sobre la pregunta 20 (Pre y Post) .....	98
Figura 46: Resultados descriptivos sobre la pregunta 21 (Pre y Post) .....	99
Figura 47: Resultados descriptivos sobre la pregunta 22 (Pre y Post) .....	100
Figura 48: Resultados descriptivos sobre la pregunta 23 (Pre y Post) .....	101
Figura 49: Resultados descriptivos sobre la pregunta 24 (Pre y Post) .....	102

## RESUMEN

La investigación se realizó con el objetivo para determinar el nivel de correlación que tiene la web scraping con la mejora de la publicación de noticias tecnológicas en el portal *web ojotecnologico*, el propósito de la conclusión de la investigación será servir de modelo para portales web que se dedican a publicar noticias.

Metodológicamente la investigación se desarrolló de tipo básico, con diseño no experimental, en el nivel descriptivo y con un resultado de significancia bilateral aceptable estadísticamente a través de las pruebas de hipótesis planteadas aplicando el instrumento de acopio de datos que utiliza la escala de Likert, la muestra fue de 66 usuarios, que son seguidores de nuestro *fanpage de ojotecnologico*.

Los resultados demuestran que las variables tienen una distribución normal, y una significancia bilateral de 0,000 menor al error estadístico aceptable de 0,05 con un nivel de significancia de 95%. Así mismo obtuvimos como resultado de fiabilidad de 0,061 el cual nos permitieron tener éxito en la investigación de saber la mejora de las publicaciones con implementar la técnica web scraping.

La investigación concluye que la implementación de la web scraping se correlaciona y significativamente con la mejora de la publicación de noticias en el portal web *ojotecnologico*, es necesario incrementar las noticias diarias, la cantidad de visitas y la usabilidad son parte de las mejoras que se ha logrado demostrar.

**Palabras clave:** Web Scraping, Noticias Tecnológicas, Aplicación Web, Automatización, Procesos, usabilidad.

## ABSTRACT

The research was carried out with the objective of determining the level of relationship that the web scraping has with the improvement of the publication of technological news in the ojotechnological web portal, the final purpose of the investigation will be to serve as a model for web portals that are dedicated to publish news

The research of basic type was developed methodologically, with non-experimental design, at the descriptive level and with a statistically acceptable result of bilateral significance through the hypothesis tests raised by applying the data collection instrument that uses the Likert scale; The sample was 66 users, who are followers of our ojotecnologico fanpage.

The results show that the variables have a normal distribution, and a bilateral significance of 0.000 lower than the acceptable statistical error of 0.05 with a level of significance of 95%. We also obtained as a result of reliability of 0.061 which allowed us to be successful in the investigation of knowing the improvement of the publications with implementing the web scraping technique.

The investigation concludes that the implementation of the web scraping is correlated and significantly with the improvement of the publication of news in the ojotecnologico web portal, it is necessary to increase the daily news, the amount of visits and usability are part of the improvements that have been managed to prove.

**Keywords:** Web Scraping, Technology News, Web Application, Automation, Processes, usability.

# I. INTRODUCCIÓN

## 1.1 Realidad problemática

Los medios digitales ya son los más consultados en todo el mundo escribe el diario el país de España en el año 2006, donde la Unión Internacional de Telecomunicaciones (UIT). En su informe Digital life 2006, difundido hoy, la organización de la ONU reflexiona sobre los cambios que ha introducido la tecnología digital en todo el mundo.

Esto dio inicio que el consumo digital de noticias haya evolucionado de lo tradicional a los computadores o mejor aún a los dispositivos móviles de todo tipo que hoy en día podemos contar, ahora los medios internacionales ven que la página web tiene gran relevancia, ya que a través de él se dan a conocer noticias que los usuarios están interesados, por tal motivo la web debe contar con un diseño visual atractivo, claro y directo que favorezca la usabilidad y el fácil acceso a los lectores.

En el Perú según el último estudio realizado por Ipsos Perú, los medios tradicionales siguen liderando la preferencia con un 55 % y los medios digitales un 45%. Dada la era digital en la que vivimos, son los dispositivos de este tipo los que han ganado, sin duda, un posicionamiento importante dentro del rubro del consumo de medios; sin embargo, la pregunta para muchos especialistas e interesados en las plataformas de comunicación es si este tipo de plataformas ha logrado superar el interés y uso de los recursos tradicionales como la televisión o la radio.

Los medios digitales que lideran en Perú son el Comercio; Diario la Republica, RPP, diario Gestión, Perú 21 en otros. En los niveles socioeconómicos A/B hay una mayor inclinación por el uso de los medios digitales como lo son las redes sociales y las webs de noticias, mientras que el NSE C/D señalan invertir más horas consumiendo plataformas tradicionales.

Ojotecnologico es un portal web de redifusión de contenido debido que se publican artículos de diferentes medios de comunicación y ahorra tiempo a las personas que buscan noticias tecnológicas en internet, y su propósito principal es que encuentres la información referente a tecnología en un solo sitio en el menor tiempo posible.



Para publicar una noticia en el portal ojotecnologico, el primer paso que tenemos que realizar es la búsqueda de fuentes de comunicación confiables y con información actualizada, para esto ingresamos a los medios de comunicación más populares o más seguidos aquí en Perú y medios internacionales, que brinden información acerca de tecnología, una vez identificado el medio de comunicación lo que hace es registrarnos en los boletines diarios o fuentes RSS para poder obtener la noticias referente a tecnología.

Este proceso es un problema debido que nos toma mucho tiempo por la cantidad de fuentes que existe en internet, a continuación, se muestra los medios de comunicación que cuenta ojotecnologico. (Ver Tabla 01)

Tabla 1  
Fuentes de Comunicación

<b>Fuente</b>	<b>País</b>	<b>Sitio Web</b>
RPP	Perú	<a href="https://rpp.pe/">https://rpp.pe/</a>
Gestión	Perú	<a href="https://gestion.pe/">https://gestion.pe/</a>
El Comercio	Perú	<a href="https://elcomercio.pe/">https://elcomercio.pe/</a>
Computer hoy	España	<a href="https://computerhoy.com/">https://computerhoy.com/</a>

Fuente propia

Actualmente las noticias tecnológicas son revisadas que cuenten mínimo con una imagen para ser publicadas en ojotecnologico, el inconveniente que tenemos es que debemos revisar noticias por noticia y hacer un filtro manual para ser descartada o si la noticia es muy relevante en cuanto a tecnología se le busca una imagen de internet que vaya acorde al tema de la noticia y de esta manera recién ser guardada en la base de datos para su publicación.

Y si la noticia cuenta con varias imágenes en su contenido se debe elegir cual de la imagen ira con la página principal con una descripción pequeña de la noticia.

Las noticias tecnológicas en algunos casos cuentan con enlaces externos, que por el momento es un problema que no podemos controlar debido a que nos llevaría mucho tiempo revisar cada noticia al detalle si cuenta o no con enlaces externos, somos conscientes que esto genera un disgusto al lector, el cual se tratará de dar solución con el método scraping.

El copiado de noticias lo hacemos de forma manual somos tres personas de que le dedicamos tiempo de acuerdo a nuestra disponibilidad, y esto ocasiona que obviemos texto al momento del copiado de la noticia y nos damos cuenta ya cuando las noticias fueron publicadas.

Tratamos en lo posible que se copie todo el texto de la noticia y no generar malestar a los lectores.

El tiempo que empleamos para guardar las noticias en la base de datos es alto debido a la cantidad de noticias por día que se pueda insertar, si eran pocas lo hacíamos manualmente, y en otros casos utilizamos un Excel para hacer una importación directa a la BD, el cual en ocasiones presentamos problemas o errores que nos retrasa en la publicación de noticias.

El horario de publicación de noticias es un problema y esto se debe que se va publicando en diferentes horas del día de acuerdo a nuestro tiempo y disponibilidad de las personas que trabajamos en ojotecnologico, no tenemos un horario fijo y eso representa un problema crítico ya que si un usuario ingresa por la mañana lo más probable que encuentre noticias del día anterior, la gran cantidad de noticia se están publicando por la tarde y noche, un punto de mejora con la web scraping es establecer un horario fijo de la publicación de noticias y de esta manera tener noticias actualizadas en tiempo real.

La necesidad de estar informados y actualizados sobre noticias tecnológicas en la mayoría de casos para los usuarios se les dificulta o es un problema porque tiene que estar buscando y abriendo una y otra página para encontrar noticias de su interés, por tal motivo se propone la mejora de publicaciones de noticias tecnológicas en el portal web ojotecnologico donde un solo sitio pueda encontrar noticias tecnológicas, y de esta manera pueden ahorrar tiempo.

En cuanto a la cantidad de noticias publicadas por día si tenemos un serio problema debido que, al tener pocas fuentes de información, ya que el trabajo de publicar lo realizamos manualmente y somos tres personas que trabajamos a diario para poder tener en promedio de 5 publicaciones por día de los cuatro medios de comunicación que actualmente contamos.

A continuación, mostramos la cantidad de noticias por día de una muestra de del mes de junio del 2021 de los primeros 15 días. (Ver Figura 1).

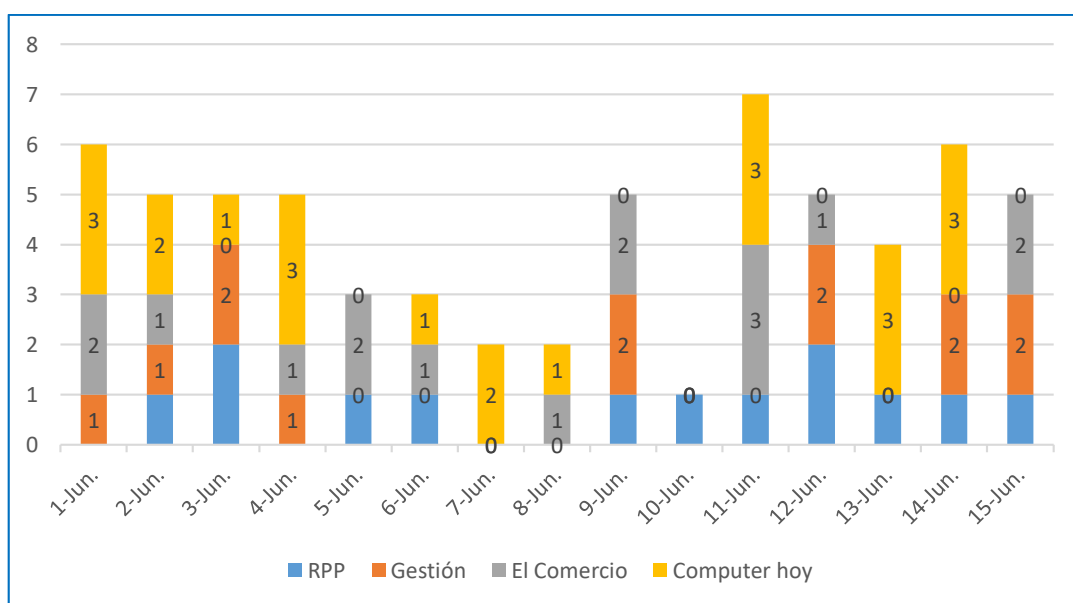


Figura 1: Cantidad de noticias publicadas por día y medio de comunicación  
Fuente: ojotecnologico  
Elaboración: propia

Podemos observar en la gráfica que hay déficit de publicar noticias diarias, en esta muestra podemos darnos cuenta que no se logra publicar por lo menos una publicación por medio de comunicación, también notamos que algunos en el 10 de junio no se llegó a publicar ninguna noticia, esto efectivamente generará insatisfacción de párate de los lectores o seguidos que diariamente vistan el portal web ojotecnologico.

Nos hay una proporcionalidad en cuanto a la cantidad de noticias publicadas diariamente de cada medio de comunicación, esto se puede deber porque en la fuente original de la noticia tampoco no publicaron noticias, también por falta de tiempo que no se llegó a guardar en la base de datos de ojotecnologico, estamos convencidos que si no se muestra

noticias diarias de cada fuente, los usuarios dejaran de visitar el portal web, y es to lo deducimos porque cada usuario tiene preferencia por uno u otro medio de comunicación.

No hay objetivos establecidos para lograr una cierta cantidad de publicaciones de cada fuente de información, que permitan hacer un seguimiento y control para que no pueda transmitir que haya favoritismos por alguna fuente de comunicación en particular.

La satisfacción se ve afectada por:

- ✓ la poca cantidad de noticias,
- ✓ fuentes de comunicación,
- ✓ noticias antiguas,
- ✓ no hay variedad de noticias de acorde lo que busca un lector,

Si bien es cierto son cuatro medios de comunicación que lo pueden visualizar en solo sitio, no es suficiente para poder mejorar la satisfacción del lector.

En una encuesta realizada se consultó sobre las satisfacciones que le ocasiona ojotecnologico y también y si volvería a ingresar en los próximos 15 días y los resultados fueron los siguientes. (Ver Figura 2).



Figura 2: ¿Cuál es el nivel de satisfacción general con ojotecnologico?

Fuente: ojotecnologico

Elaboración: propia

La realidad de ojotecnologico es que existe un 47% que visita el portal web y se siente insatisfecho, quizás porque no encontró lo que buscaba, pocas noticias publicadas, no encontró su medio de comunicación favorito, a través de estos resultados es que se plantea mejorar la cantidad de noticias aplicando la web scraping y así de esta manera poder incrementar las noticias y las fuentes de comunicación y de este modo revertir el alto índice de insatisfacción. (Ver Figura 3).



Figura 3: ¿Cree que volverá a visitarnos en los próximos 15 días?

Fuente: ojotecnologico

Elaboración: propia

Vemos que el porcentaje de usuario que no ingresarían en un lapso de 15 días es alto, esto nos da entender que la web no cumple con sus expectativas y que se tiene que plantear mejorar para poder revertir la situación, por el otro lado el 53% que si ingresaría hay un riesgo alto que ya no ingresen por falta de contenido relevante o aumentar las fuentes de información y que no haya noticias actualizadas.

Otro problema que afecta a la satisfacción del lector es que la web no esté disponible y esto se debe que el servidor este caído, que está sucediendo muy recurrente, para solucionar de manera rápida se tiene que hacer un reinicio, por lo cual se debe tomar la decisión de migrar de servidor que si garantice que la web esté disponible las 24 horas y los 365 días del año.

La propuesta planteada en la investigación es de resolver los procesos de publicación de noticias el cual minimiza en tiempo y costo de publicación, a su vez mejorar el índice

de satisfacción de los lectores ya que pudran encontrar noticias actualizadas de diferentes medios de comunicación digital ya sea nacional o internacional.

Empleando práctica scraping estamos seguros que permitirá automatizar de manera oportuna de brindar las noticias tecnológicas a los usuarios, cabe indicar que mejorar los beneficios para el portal web ojotecnologico, ya que podría fidelizar lectores y tener recomendación hacia la web.

## **1.2 Planteamiento del problema**

### **1.2.1 Problema general**

¿De qué manera la implementación de una aplicación web scraping mejorará la publicación de noticias tecnológicas en el portal *web ojotecnologico*?

### **1.2.2 Problemas Específicos**

- a) ¿En qué medida la cantidad de noticias publicadas por día de manera óptima permitirá la mejora de las publicaciones de noticias en la *web ojotecnológico*?
- b) ¿En qué medida la cantidad de visitas por día permitirá mostrar la afluencia de usuarios y a su vez la mejora de las publicaciones de noticias en la *web ojotecnológico*?
- c) ¿En qué medida el nivel de usabilidad del aplicativo web permitirá facilitar el manejo a los usuarios y así mismo la mejora de las publicaciones de noticias en la *web ojotecnológico*?

### **1.2.3 Delimitación del problema de investigación Espacial**

Esta investigación recopilará y analizará la información referente al problema de proceso de publicación de noticias en el portal ojotecnologico, que actualmente se puede visitar ingresando a la url [www.ojotecnologico.com](http://www.ojotecnologico.com).

#### **1.2.4 Delimitación del problema de investigación Temporal**

El objeto de la investigación tomará como punto de partida el mes de abril a diciembre del dos mil diecisiete (2021), se considera un periodo que permite establecer los objetivos planteados.

### **1.3 Hipótesis de la investigación**

#### **1.3.1 Hipótesis General**

La Implementación de una aplicación Web Scraping mejora significativamente la publicación de noticias tecnológicas en el portal *Web ojoTecnológico*.

#### **1.3.2 Hipótesis Específicas**

- a) El Aumento de la cantidad de noticias publicadas por día, mejora y automatiza la publicación de noticias en la *web ojotecnológico*.
- b) El Incremento de la cantidad de visitas por día demuestra la afluencia de usuarios y a su vez mejora la publicación de noticias en la *web ojotecnológico*.
- c) Si se garantiza el nivel de usabilidad entonces se facilita el manejo del aplicativo por parte de los usuarios y así mismo se mejora la publicación de noticias en la *web ojotecnológico*.

### **1.4 Objetivos de la investigación**

#### **1.4.1 Objetivo General**

Implementar una aplicación Web Scraping, para mejorar la publicación de noticias tecnológicas en el portal *web ojoTecnológico*.

### 1.4.2 Objetivos Específicos

- a) Aumentar la cantidad de noticias publicadas por día para mejorar y automatizar las publicaciones de noticias en la *web ojotecnológico*.
- b) Aumentar la cantidad de visitas por día para mostrar la afluencia de usuarios y a su vez la mejora de las publicaciones de noticias en la *web ojotecnológico*.
- c) Garantizar el nivel de usabilidad para facilitar el manejo del aplicativo por parte de los usuarios y así mismo la mejora de las publicaciones de noticias en la *web ojotecnológico*.

### 1.5 Variables, dimensiones e indicadores

A continuación, en tabla (2) se detalla las variables de la investigación con sus dimensiones e indicadores que nos permitirá demostrar que la aplicación de la web scraping ha ayudado con la mejora de publicación de noticias.

Tabla 2  
Variables, dimensiones e indicadores.

Variables	Dimensiones	Indicadores
Aplicación Web Scraping	<ul style="list-style-type: none"> <li>• Noticias por día publicadas.</li> <li>• Visitas por día.</li> <li>• Usabilidad.</li> </ul>	<ul style="list-style-type: none"> <li>• Cantidad de noticias publicadas por día.</li> <li>• Cantidad de visitas por día.</li> <li>• Nivel de usabilidad.</li> </ul>
Publicación de noticias.	<ul style="list-style-type: none"> <li>• Satisfacción</li> <li>• Costos</li> <li>• Calidad de servicio.</li> </ul>	<ul style="list-style-type: none"> <li>• Nivel de satisfacción.</li> <li>• Reducciones de costos.</li> <li>• Nivel de calidad</li> </ul>

Fuente propia



## **1.6 Justificación del estudio.**

### **1.6.1 Justificación Teórica**

Con el estudio que se realizará en esta tesis para la implementación de la web scraping que permitirá ser usado con fines de publicar contenido web, lo que lograría optimizar los procesos de búsqueda de información y ponerlo a disponibilidad de los usuarios que navegan por el internet. Es una manera de justificar que la web scraping es una buena práctica para la obtención de información respetando la autoría y fuentes de origen.

### **1.6.2 Justificación Práctica**

La implantación de la tesis dará resultados concretos que permiten solucionar y optimizar el proceso de la publicación, de noticias, ayudaría contar con más fuentes de información y de esta manera mejoraría la satisfacción de los lectores,

### **1.6.3 Justificación Metodológica**

El tipo de investigación que utilizaré para mi trabajo de tesis es la siguiente. La investigación aplicada, utilización de los conocimientos en la práctica, para aplicarlos en provecho de los grupos que participan en esos procesos y en la sociedad en general, además del bagaje de nuevos conocimientos que enriquecen la disciplina. El nivel de investigación que se utilizará en el desarrollo de la de tesis es descriptiva el cual tiene como objetivo la descripción de los fenómenos a investigar, tal como es y cómo se manifiesta en el momento (presente) de realizarse el estudio y utiliza la observación como método descriptivo, buscando especificar las propiedades importantes para medir y evaluar aspectos, dimensiones o componentes.

En cuanto al enfoque de la investigación será cuantitativo: Representa, un conjunto de procesos, es secuencial y probatorio.

#### **1.6.4 Justificación Social**

La implantación de la tesis tendrá un efecto positivo hacia la sociedad, permitirá el libre acceso a la información y generará valor hacia el lector que incrementará sus conocimientos en cuanto a tecnología.

#### **1.6.5 Justificación Legal**

El objeto de estudio de la tesis se realiza cumpliendo con las normas y políticas de las fuentes de información. Se tomará la precaución de los datos personales serán tratados con el grado de protección adecuado, tomándose las medidas de seguridad necesarias para evitar su alteración, pérdida, tratamiento o acceso no autorizado por parte de terceros. De acuerdo con lo establecido en la Ley 29733 (Ley de Protección de Datos Personales y su Reglamento).

#### **1.6.6 Justificación Económica**

Una vez implementado la tesis de investigación, permitirá a los usuarios a ahorrar tiempo en la búsqueda de noticias tecnológicas y de esta manera podrá estar informando de los avances de la tecnología.

#### **1.6.7 Importancia**

Esta investigación servirá para mejorar el proceso de publicación de las noticias tecnológicas, el cual servirá como modelo para futuras investigaciones que deseen usar o implantar la técnica scraping para mejorar sus procesos.

La utilidad que tendrá esta investigación para poder demostrar que se puede recolectar información de importancia de cualquier medio de comunicación sin violar los derechos de autor, para estar informados en tiempo real de cualquier fuente de información.

Nuestro proceso de scraping es capaz de procesar grandes volúmenes de información en un tiempo que podríamos decir relativamente corto el cual está provisto con sólidas medidas de seguridad cuando se trabaja con información sensible como sería la protección de datos personales, otra medida que optamos

es que continuamente se pueden hacer cambios estructurales para los procesos de publicación de noticias el cual es prescindible para la mejora continua.

A nivel empresarial veremos la utilidad que tiene la web scraping según (Fiumara & Baumgartner, 2014), indica que sistema de extracción de datos Web es una plataforma que implementa una secuencia de procedimientos que extraen información de orígenes Web, asimismo proponen la división de dichos sistemas en dos niveles, nivel empresarial y nivel social.

Las aplicaciones del nivel empresarial son:

- ✓ Publicidad de Contenido Consciente.
- ✓ Atención al Cliente.
- ✓ Construcción de Bases de Datos.
- ✓ Ingeniería de Software.
- ✓ Pruebas Funcionales de Aplicaciones Web
- ✓ Comercio Comparativo
- ✓ Escenarios de Aplicaciones Web Híbridas
- ✓ Minería de Opinión y Análisis de Sensibilidad
- ✓ Bases de Datos de Citas Bibliográficas

(Cenys & Grigalis, 2013) Nos comenta que la inteligencia de negocios en línea, que incluye extracción, integración, análisis, y distribuyendo información sobre productos, clientes, competidores, etc., ayuda empresas para tomar mejores decisiones y lograr una ventaja competitiva.

Sin embargo, la extracción de datos de muchas fuentes web es un proceso que requiere esfuerzo humano que trae costes considerables para la organización muchos avances tecnológicos prometedores y direcciones de investigación están tratando de totalmente automatizar la extracción de datos web y por lo tanto ayudar a las organizaciones a minimizar los costos.

Negocios en línea especialistas en inteligencia en organizaciones actualmente se puede confiar en la web asistida visualmente sistemas de extracción de datos que

minimizan los esfuerzos humanos necesarios para extraer datos de los sitios web de destino.

A continuación, se lista en un cuadro de las webs comerciales que utilizan la técnica de extracción de información (Ver tabla 3)

Esta investigación será muy útil para el problema de que el portal ojotecnologico no mostraba un numero de noticias diarias y actualmente se puede indicar que ingresa un promedio de diez (10) noticias por cada medio digital y esto es aleatorio según el medio de comunicación, el sistema web scraping está bien programada para que detecte y jale las noticias que son publicadas por los medios de comunicación aquí en Perú como los medios extranjeros.

Tabla 3  
Sistemas de extracción de información web disponibles comercialmente

#	Nombre	Produc / Serv	Descarga inmediata	Sitio web	Precio
1	Automation Anywhere	Producto	Si	www. automationanywhere. com	Desde \$955
2	Connotate	Servicio.	No	www. connotate.com	N/A
3	Denodo	Servicio	No	www. denodo.com	N/A
4	Djuggler	Producto.	Si	www. djuggler.com	Desde \$249
5	iOpus iMacros	Producto	Si	www. iopus.com	Desde \$495
6	Kapow	Servicio	Si	www .kapowsoftware.com	N/A
7	Lixio	Servicio	No	www. lixto.com	N/A
8	Mozenda	Servicio	Si	www. mozenda.com	Desde \$299/mes
9	Visual Web Riper	Producto	Si	www. visualwebripper.com	Desde \$299
10	WebSundew	Producto	Si	www. websundew.com	Desde \$69

Fuente: Cenys & Grigalis  
Elaboración: propia

Otro problema que será solucionado son la noticias que no cuentan con una imagen o que cuentan con galerías de imágenes, ahora con el sistema web scraping podemos filtrar y controlar las noticias que cumplan ciertos criterios establecidos por el portal ojotecnologico de esta manera el usuario tendrá una

navegación por la web sin que se encuentre con alguno error o una noticia incompleta

Con respecto a la técnica web scraping que se utiliza y que se implementó en el portal ojotecnologico vemos que en la actualidad tiene mucho auge en cuanto a la innovación y su uso tal como lo indica el medio español computing.

Las tendencias para este año según investigaciones son: Web Scraping, Machine Learning y Blockchain aplicado a la ciberseguridad.

La digitalización de las empresas a nivel nacional e internacional ya un hecho: Big Data, ciberseguridad, IA ordena los planes y estrategias de las empresas para los próximos meses. Por ello, Vector ITC Group desvela alguna de las tendencias más destacadas para 2020, el objetivo de asistir a las distintas empresas a lograr el mejor beneficio a su transformación digital.

Web scraping es una técnica practica usada mediante programas de software para extraer información de sitios web. Generalmente, estos programas emulan la navegación de un humano en internet, ya sea utilizando el protocolo HTTP, o embebido en un navegador de una aplicación. Son instrumentos que forman parte de las tecnologías Big Data y Analytics y se caracteriza por su capacidad para sacar información pública.

La tecnológica tiene un fin de mejorar y optimizar las formar de recolectar y procesar información el cual le da un valor importante para que esta pueda permitir tomar decisiones en diferentes investigaciones ya sea publica o privadas.

Los beneficios a nuestros lectores de ojotecnologico es mantenerlos informados de las noticias tecnológicas de diferentes medios de comunicación en tiempo real, esta investigación permite que los lectores encuentren gran cantidad de noticias según sus gustos y preferencias y sobretodo el ahorro de tiempo significativo para estar informado.

En cuanto a los beneficios a la sociedad que se aporta con nuestra investigación es que por un lado cuando se usan la tecnología adecuadamente se brinda soluciones para el fácil acceso a la información al cual todos tenemos derecho de estar informados, y por otro lado podemos ser tomados como modelo para futuros emprendimientos y dar soluciones a necesidades que aún falta satisfacer a los usuarios.

Esta investigación abrirá nuevas opciones de estudios sustantivos que presenten situaciones similares a lo que aquí se está planteando, puede ser tomado como marco de referencia.

## 1.7 Trabajos Previos.

### **Título: Web Scraping Para La Obtención De Información Actualizada En Internet Con Push Notifications Para Smartphone**

Tesis para optar el título de licenciatura informática que presenta

Autor: Hugo Alfredo Condori Yujra

Centro de Estudio: Universidad Mayor De San Andrés.

Ciudad / País: La Paz Bolivia – 2014

Fecha de captura: 09 de septiembre 2021.

URL: <http://repositorio.umsa.bo/bitstream/handle/123456789/8405/T.2874.pdf>

Web Scraping (raspado web o extracción de datos web) es el proceso de extracción de datos de sitios web, preferiblemente usando un programa que simula la exploración humana mediante el envío de peticiones HTTP simples o emulando un navegador web completo. Web Scraping, Content Scraping, Screen Scraping, Web Harvesting o Web Data Extraction son todos términos análogos.

En general, cualquier cosa que se puede ver en Internet puede ser extraído y este proceso puede ser automatizado. Web Scraping se utiliza normalmente por diferentes motivos, como la detección de cambio, la investigación de mercado, seguimiento de datos y en algunos casos hasta el robo de datos. Web Scraping está muy relacionado con la indexación de la web, la cual indexa información de

la web utilizando un agente web automatizado y es una técnica global adoptada por la mayoría de los motores de búsqueda.

Pero, Web Scraping se enfoca más en la transformación de datos sin estructura en la web, en datos estructurados que pueden ser almacenados y analizados en una base de datos central, en una hoja de cálculo o en alguna otra fuente de almacenamiento.

Existe una polémica con aspectos legales por la obtención de los datos, no tanto por la forma, si no de lo que se va hacer después con los datos, por ejemplo, no sería legal si muestra datos privados o con copyright.

Funcionamiento de Web Scraping.

El funcionamiento de Web Scraping consiste en una simulación de la navegación, que una persona realizaría, ya sea implementado a bajo nivel, como en HTTP (Hypertext Transfer Protocol), o incluido en ciertos navegadores web.

Ésta simulación se realiza mediante programas denominados “bots”, que crean una clonación del cliqueo, la lectura y el “coppypaste”, automatizando la tarea de búsqueda y de recolección de datos.

De modo que, Web Scraping se centra en la transformación del contenido no estructurado, por lo general en formato HTML, en datos estructurados que pueden ser almacenados y analizados.

El uso de Web Scraping para empresas o personas son muchos, ya que cada empresa o individuo tiene su propia necesidad específica de recolección de datos. Algunos de estos usos son los siguientes.

- ✓ Recolección de noticias, artículos, blogs, etc.
- ✓ Periodismo de datos.
- ✓ Monitoreo a artículos y datos cambiantes.
- ✓ Comparación de precios en línea.
- ✓ Recolección contactos.
- ✓ Investigación.

Existen diversos factores que dificultan la obtención de los datos mediante el uso de herramientas destinadas al web scraping, estos factores pueden hacer que realizar la obtención sea más complicado o incluso llegando a impedir que se realice. Algunos de estos factores son: Ejemplos de dispositivos móviles en las que se encuentra esta tecnología:

Códigos HTML mal estructurados, por ejemplo, aquellas páginas que fueron creadas hace tiempo y no se han adaptado, de modo que su estructura es muy simple y no es posible extraer de su etiquetado HTML información estructurada.

Páginas que contiene sistemas de autenticación como códigos y paywalls de CAPTCHA ( Completely Automated Public Turing test to tell Computers and Humans Apart). Resulta evidente que conseguir la información mediante la programación de un scraper es imposible, ya que para acceder a la página es necesario realizar una autenticación de CAPTCHA que es totalmente diferente en cada petición.

Bloqueo al acceso masivo por los administradores de los servidores. Al programar un “scraper” se realizan muchas peticiones a un servidor, el cual puede detectarlo como un ataque de denegación de servicio y bloquear nuestra IP.

Sistemas que usan cookies para realizar un seguimiento de lo que el usuario realiza.(Yujra & Alfredo, 2014)

**Título: Método general de extracción de información basado en el uso de lógica borrosa. Aplicación en portales web**

Tesis doctoral que presenta

Autor: Ropero Rodríguez, Jorge

Centro de Estudio: Universidad de Sevilla. Departamento de Tecnología Electrónica.

Ciudad /Pais: Sevilla -2009

Fecha de captura: 09 de septiembre 2021.

URL: <https://www.dte.us.es/personal/jropero/TesisV2p0.pdf>



La gran cantidad de información disponible en la actualidad provocada por el auge de las Tecnologías de la Información constituye una enorme ventaja para las necesidades de búsqueda de esta por parte de los usuarios de las nuevas tecnologías.

Sin embargo, al mismo tiempo, surge también un gran problema derivado de la dificultad existente para distinguir la información necesaria de entre toda la enorme cantidad de datos innecesarios. Por esta razón, los conceptos de Búsqueda de Información (Information Retrieval, IR) y Extracción de Información (Information Extraction, IE) han saltado a la palestra con fuerza en los últimos tiempos.

En principio, ambas surgieron para la búsqueda y extracción de documentos, pero en los últimos años se ha generalizado su uso para la búsqueda de cualquier otro tipo de información, como puede ser la información contenida en una base de datos, una página web o, en general, cualquier conjunto de conocimiento.

En particular, está muy extendido el uso del denominado Modelo de Espacio Vectorial (Vector Space Model, VSM), el cual está basado en la utilización de términos índice a los que se asocia un cierto peso, que representa la importancia de estos en el conjunto de conocimiento a considerar. Estas técnicas funcionan razonablemente bien a la hora de extraer información en muchos ámbitos, pero poseen el inconveniente de no ser tan eficientes cuando las búsquedas de un usuario no son demasiado concretas, o bien cuando existe una ingente cantidad heterogénea de información.

Por este motivo, se propone en esta tesis doctoral el desarrollo de un Agente Inteligente (también denominado en la bibliografía Asistente Virtual) que sea capaz de responder a las necesidades de los usuarios en su proceso de encontrar la información deseada en entornos en los que la información es ingente, heterogénea, vaga, imprecisa o desordenada.

La aportación principal de esta tesis es la creación de un método general para la búsqueda y extracción de información basado en el uso de la Lógica Borrosa (Fuzzy Logic, FL), la cual es una herramienta ideal para la gestión de una información de las características antes mencionadas, y, en particular, la aplicación y validación de este método para la extracción de información de portales web, dado que los portales web son un claro exponente de información heterogénea y desordenada. (Ropero Rodríguez, 2010)

**Título: Sistema de extracción de información basado en ontologías para comentarios de un foro de discusión en línea en el dominio de cursos brindados por una entidad de educación superior.**

Tesis para optar el título de Licenciado en Ingeniería Informática

Autor: Peña Vilca, Willy Alexis

Centro de Estudio: Pontificia Universidad Católica del Perú.

Ciudad /Pais: Lima/Perú 2015

Fecha de captura: 17 de septiembre 2021.

URL: <http://tesis.pucp.edu.pe/repositorio/handle/123456789/6064>

En la actualidad, las empresas necesitan estar informadas acerca de la opinión que tienen sus principales clientes respecto a los productos o servicios que ofrecen.

Esto se debe a que estas dependen de esa información para poder tomar decisiones estratégicas al respecto. Para poder lograr esto, muchas de ellas optan por contratar servicios de empresas consultoras que realicen una encuesta tradicional o un focus group para poder obtener la información requerida; sin embargo, la parcialidad que guardan este tipo de estudios hace que en algunos casos se tomen decisiones estratégicas a partir de información no del todo fiable o representativa.

En base a lo mencionado anteriormente, el presente proyecto de fin de carrera brindará una herramienta para poder aprovechar la información contenida dentro de los comentarios hechos en foros de discusión en línea. Estas fuentes de

conocimiento muchas veces no son procesadas para ningún fin; sin embargo, por medio de la herramienta propuesta se podrá extraer información relevante para ser utilizada como base de conocimiento por una empresa del rubro educativo.

A partir de la cual se podrá contar con una alternativa confiable para obtener información sobre la opinión directa de sus alumnos respecto a los cursos y profesores pertenecientes a la organización. Por último, dicha información podrá ser utilizada como base durante la toma de decisiones estratégicas de dicha organización educativa.

En este proyecto se presenta un enfoque para la extracción automática de los datos del autor, la fecha y el contenido de una publicación realizada en un foro de discusión. Los resultados del experimento muestran que el enfoque es efectivo; sin embargo, no puede ser utilizado en páginas web que generan dinámicamente datos a través de JavaScript.

En este proyecto, se utiliza un método que utiliza el algoritmo de alineación de árboles y el método de transferencia de aprendizaje, los cuales son propuestos para generar las capas de páginas web de foros de discusión en línea, blogs y páginas de noticias. Para tal fin se consideran los siguientes puntos:

El algoritmo de alineación de árboles es utilizado para encontrar la estructura más similar de las páginas web usadas. Luego, se generan regresiones lineales para darle un peso a dichas comparaciones. Basado en esto, se mezcla los árboles en uno solo que almacena información estadística obtenida de múltiples páginas web.

El método de transferencia de aprendizaje es utilizado para obtener el contenido de bloque más parecido y aplicar el algoritmo mencionado anteriormente para encontrar patrones repetidos en el árbol encontrado.

Por último el método de generación de capas mide la similitud entre las capas y las páginas web; en caso se encontrará algún cambio drástico en aquella similitud se genera una nueva capa según datos estadísticos.(Vilca & Alexis, 2015) .

**Título: Medición de la satisfacción del cliente y el impacto de la publicidad en unidades productivas de universidades públicas peruanas**

Tesis para para optar el grado académico de maestro en ciencias

Autor: Flores Bashi, Carlos Antonio

Centro de Estudio: Universidad Nacional de Ingeniería

Ciudad / País: Lima- Perú – 2012

Fecha de captura: 10 de octubre 2021.

URL: <http://cybertesis.uni.edu.pe/handle/uni/3324>

### **LA SATISFACCIÓN DEL CLIENTE**

Al respecto, se observa opiniones diversas mientras algunos hacen énfasis en altos niveles de satisfacción para conseguir altos índices de rentabilidad. Otros remarcan lo contrario y que estos altos niveles de satisfacción no necesariamente guardan relación con la rentabilidad del negocio, a continuación, se detalla algunas de estas propuestas.

De Fernando Gosso (2008) saco una conclusión que todo servicio o producto debe brindar a sus clientes es que se les debe mantener Hipersatificados y esto debe ir en coherencia con un buen servicio brindado y a cambio recibir un agradecimiento del cliente y por ende esto ayudará a crecer el negocio porque habrá una buena recomendación.

Luego de revisar la información que proporciona Kotler & Lane y Claes Fornell (2008), coinciden en algo muy relevante y que se debe tomar en cuenta que la medir los indicadores de satisfacción con los indicadores de ingresos y beneficios no hay una correlación ya que hablan en sus estudios donde hubo índices de satisfacción muy baja sim embargo obtuvieron buenas ganancias, en cual nos lleva a reflexionar que es lo que ocasiona esta situación que los servicios o productos son necesarios para los usuarios que no lo toma el interés debido de cómo fue atendido por un determinada empresa, podemos hacer un conclusión que una alta satisfacción no nos asegura que tengas buenos ingresos por un producto o servicio.

Las diferentes industrias hoy en día plantean diferentes variables para poder medir la satisfacción y se están basando a una satisfacción de producto, servicio y apelan a la sociedad el que puede permitir tener un resultado más fidedigno y satisfactorio.

### ✓ **Tipos de Indicadores.**

Se pueden clasificar los indicadores en indicadores de eficacia o de eficiencia. El indicador de eficacia mide el logro de los resultados propuestos. Indica si se hicieron las cosas que se debían hacer, si se está trabajando en los aspectos correctos del proceso.

Los indicadores de eficacia se enfocan en el que se debe hacer, por tal motivo en el establecimiento de un indicador de eficacia es fundamental conocer y definir operacionalmente los requerimientos del cliente del proceso para comparar lo que entrega el proceso contra lo que la espera.

Este resultado indica que los clientes no están satisfechos, que su grado de satisfacción es bajo y, por lo tanto, se están realizando actividades incorrectas que no agregan valor a los usuarios del proceso. Pero es posible que el proceso sea muy eficiente entregándole al usuario información al instante que no quiere.

Los indicadores de eficiencia miden el nivel de ejecución del proceso, se concentran en el cómo se hicieron las cosas y miden el rendimiento de los recursos utilizados por el proceso tienen que ver con la productividad.

La importancia de los indicadores es que nos pueden permitir tomar acciones inmediatas para corregir o cambiar de estrategia de cara la satisfacción de nuestro producto o servicio brindado, y tiene mucho que ver con los procesos definidos de cómo se realice cada procedimiento al momento de brindar un servicio o vender un producto.

✓ **Medición de la satisfacción del cliente.**

La medición de la satisfacción del cliente es importante para la comparación y control del nivel de satisfacción de los clientes, de tal forma que justifique asignar recursos financieros para este fin Shankar y Carpenter (2012).

Lo que realiza ACSI es muy interesante para la medición de la satisfacción basado en entrevistas a clientes de diferentes industrias y sectores económicos tanto de servicios como de productos. El cual nos detallan de cómo se realiza

Señala tres pasos para la formulación y uso del cuestionario:

- ✓ Paso 1: Identificar las necesidades de los clientes o las dimensiones de la calidad, se proponen ejemplos específicos de tales dimensiones.
- ✓ Paso 2: Diseñar y evaluar el cuestionario, debe corresponder a las necesidades básicas de los clientes, identificadas en el paso 1.
- ✓ Paso 3: Representa los diversos usos específicos de los cuestionarios para evaluar la satisfacción del cliente, esto puede corresponder desde identificar la situación actual de la satisfacción del cliente hasta evaluarla en el transcurso del tiempo. (Ver figura 4)

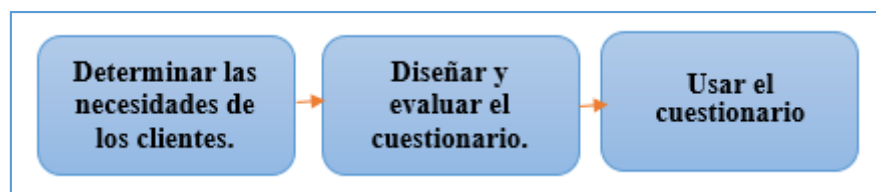


Figura 4: Modelo general para la formulación y uso de cuestionarios  
 Fuente: Bob E. Hayes "Como medir la satisfacción del cliente" 2008  
 Elaboración: propia

### **Elemento de satisfacción**

Después de formar los grupos, se redacta una frase para cada uno de ellos que refleje el contenido de sus incidentes. Esta frase se denomina elemento de satisfacción.(Bashi & Antonio, 2012)

## 1.8 Teorías relacionadas al tema

### 1) Procesos.

Según los autores (León, Rivera, & Nariño, 2009), se entiende por proceso la:

“Secuencia ordenada de actividades repetitivas que se realizan en la organización por una persona, grupo o departamento, con la capacidad de transformar unas entradas (Inputs) en salidas o resultados programados (outputs) para un destinatario (dentro o fuera de la empresa que lo ha solicitado y que son los clientes de cada proceso) con un valor agregado.

Los procesos, generalmente, cruzan repetidamente la barrera funcional, fuerzan a la cooperación y crean una cultura de la empresa distinta (más abierta, menos jerárquica, más orientada a obtener resultados que a mantener privilegios)”.

La aplicación de la gestión por procesos en una organización permite:

- ✓ comprender la configuración de los procesos del negocio, sus fortalezas y debilidades
- ✓ determinar los procesos que necesitan ser mejorados o rediseñados
- ✓ establecer prioridades, iniciar y mantener planes de mejora que permitan alcanzar los objetivos establecidos
- ✓ reducir la variabilidad innecesaria que aparece habitualmente cuando se producen o prestan determinados servicios
- ✓ eliminar las ineficiencias asociadas a la repetitividad de las acciones o actividades, al consumo inapropiado de recursos, entre otras.

Se puede representar el concepto según la figura 5.

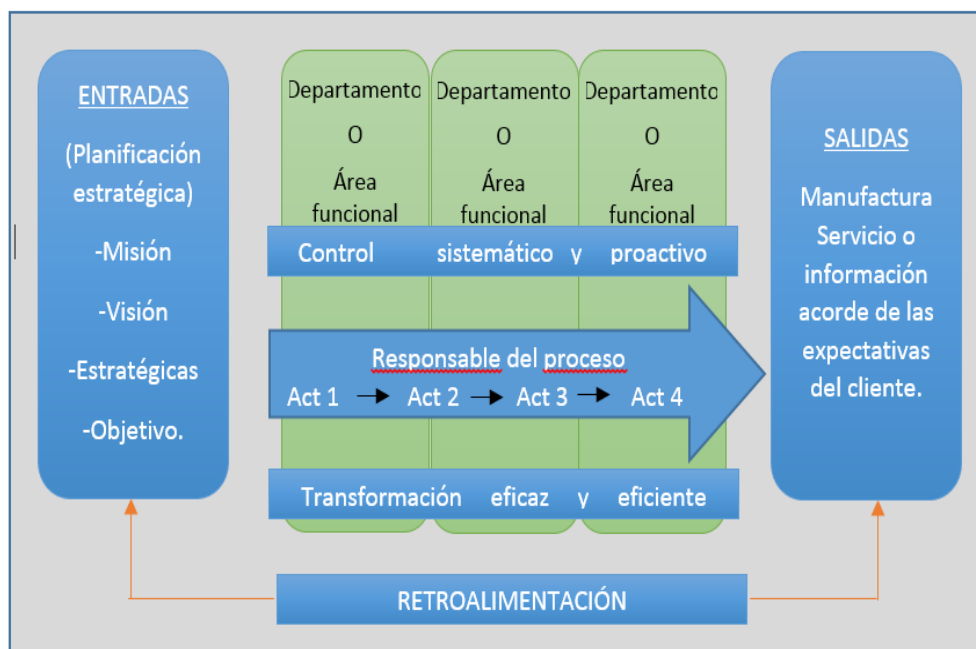


Figura 5: Representación gráfica del concepto de proceso  
Fuente: Libro- Metodología de la investigación, Quinta edición, Capítulo 1

Las principales características relacionadas para el concepto de proceso son:

- Se pueden describir las entradas y las salidas.
- Constituido por actividades internas que de forma coordinada logran un valor apreciado por el destinatario.
- Realizado por personas, grupos o departamentos de la organización,
- Cruzan uno o varios límites organizativos funcionales.
- Se habla de metas y fines, en vez de acciones y medios. Un proceso responde a la pregunta ¿Qué?, no al ¿Cómo?
- Debe poseer un responsable y ser administrado según el ciclo de Deming,
- Ser fácilmente comprendido por cualquiera,
- Poseer indicadores que visualicen de forma gráfica la evolución del mismo,
- Dinámicos; dependen de los recursos, la habilidad y la motivación del personal involucrado para generar el resultado deseado.(León, Rivera, & Nariño, 2009.)



## 2) Noticia

En el artículo publicado por RTVE.ES, en el 2010, habla sobre de que trata una noticia y de como entender cono esta estructurado y lo detalla de la siguiente manera:

Por noticia entenderemos cualquier documento informativo publicado en la web, con independencia del género periodístico al que pertenezca.

Los párrafos de la noticia deben ser cortos y habrá que evitar en lo posible las informaciones muy extensas. Es preferible dividir la información en varias capas de profundidad a las que el usuario pueda ir accediendo si está interesado.

Es mejor abusar del punto y aparte que de las oraciones subordinadas.

La noticia se crea habitualmente para su publicación inmediata y se refiere por regla general a asuntos que acaban de suceder e incluso muchas veces ni siquiera han concluido.

Usaremos siempre los tiempos verbales que más acerquen la historia al lector. Por regla general, el pretérito perfecto y el presente de indicativo. Se combinarán con los tiempos que correspondan en el caso de que las noticias que se acaban de producir se refieran a asuntos pasados o futuros.

Ejemplos:

Sí: “El accidente ha colapsado las vías de acceso a la capital, que siguen cortadas”.

No: “El accidente colapsó las vías de acceso a la capital, que siguen cortadas”

Sí: “Renfe suspende su compromiso de puntualidad a causa del temporal que azota la península. Las vías continúan colapsadas por la nieve”

No: “Renfe suspendió su compromiso de puntualidad a causa del temporal que azota la península. Las vías continúan colapsadas por la nieve.”

✓ **Elementos de la noticia:**

**Titular:** Es el elemento más importante de la noticia, pues gran parte de los usuarios se contentarán con su lectura. El titular contiene los datos básicos de la información. Ha de ser lo suficientemente atractivo para suscitar el interés del usuario y lo bastante descriptivo como para que los buscadores de internet puedan encontrarlo.

**Las claves:** Las claves o “bolitos” resumen los datos más importantes. Sirven para indicar las novedades que se han producido. Son de una línea y no llevan punto y final. Su número dependerá del diseño, pero no se abusará de ellas.

**Entradilla:** La entrada debe desarrollar el titular y completar la tesis que plantea con los datos. Las entradas no pueden limitarse a repetir el titular sin aportar más información y deben invitar al usuario a seguir leyendo.

**Cuerpo de la noticia:** El cuerpo de la noticia sirve para completar la información. Se deberá incorporar al cuerpo de la noticia un párrafo que la ponga en contexto y ayude a comprenderla.

**Firma:** Los textos que se publiquen e deberán ir siempre firmados, pues implica atribución de responsabilidades. La firma será de una agencia cuando la información que proporcione se publique sin edición. Cuando el redactor haya empleado información de agencias u otro medio para elaborar su noticia se hará constar en el texto.

**Elementos visuales:** La información debe apoyarse en elementos gráficos que guíen al usuario en la lectura. Así, el periodista podrá servirse de las negritas, los ladillos y las citas textuales u otros elementos a su alcance para facilitar el “escaneo” de la página en la que muchas veces se convierte la lectura en internet. (RTVE.ES, 2010)

### 3) Satisfacción del usuario.

Patricia Hernández Salazar(2011), autora del artículo la importancia de la satisfacción del usuario para la Universidad Complutense de Madrid nos menciona que cada vez crece el interés por saber cómo se sienten los usuarios hacia un servicio brindado y lo detalla de la siguiente manera:

Un usuario es la persona o grupo de personas que tiene una necesidad de información y que utiliza o utilizará recursos o servicios de información para cubrirla, dicho así pareciera que es muy claro el objetivo de los profesionales de la información: planear, diseñar y proveer recursos y servicios de información.

Hasta aquí estaremos realizando nuestra actividad, sin embargo, ¿nuestra actividad profesional representa nuestra esencia como disciplina?

La respuesta es que sólo representa una parte, el hacer de nuestra profesión, falta precisar la esencia, el ser de nuestra actividad, es decir, el sentido de nuestro hacer.

Para llegar a él es menester precisar que no sólo haremos, nuestro hacer debe cubrir las expectativas y necesidades de ese sujeto que necesita, utiliza o utilizará la información, concretamente tendremos que satisfacer cabalmente sus necesidades.

En su sentido más primigenio, satisfacer significar complacer un deseo, y por derivación satisfacción se entiende como la acción de satisfacer una necesidad o un deseo, el sentido de estos dos términos determina que nuestra esencia u ontología es complacer las necesidades de nuestros usuarios.

El propósito del presente trabajo es establecer la importancia que tienen los estudios de satisfacción dentro de las actividades de información, no basta con almacenar y poner a disposición de nuestras comunidades los recursos, herramientas y servicios, se hace imprescindible valorar si el suministro de dichos recursos es el adecuado para cubrir sus necesidades.

Este estado siempre es un constructor y juicio de evaluación, ya que se comparan los objetivos y expectativas contra los resultados obtenidos.

La satisfacción contempla tres grandes áreas: el funcionamiento del sistema de búsqueda y recuperación de información; el recurso de información obtenido; y los servicios.

Desafortunadamente la mayoría de estudios sobre la satisfacción del usuario se ha dirigido a medir el funcionamiento de los sistemas de búsqueda y recuperación de información y el recurso obtenido y no a los servicios.

Las medidas más utilizadas han sido la relevancia y la pertinencia, dejando de fuera elementos tales como, el estilo o diseño de la interfase, la facilidad del manejo del sistema, la utilidad de la información y los servicios de toda la unidad de información.

Algunas actitudes de los usuarios que nos muestran que tienen un descontento potencial son:

- ✓ elogiar a otras unidades de información o algún recurso o servicio específico para menospreciar otro;
- ✓ hacer una sugerencia racional;
- ✓ mostrar sus sentimientos mediante alguna conversación, comparando sus colecciones privadas como mejores que las de la biblioteca; no asistir a la biblioteca;
- ✓ presentar conductas inconvenientes (hablar fuerte, emplear palabras ofensivas).

Si percibimos alguno de estos comportamientos significa que el usuario no está satisfecho pero teme por alguna razón presentar su queja o descontento. (Patricia Hernández Salazar, 2011)

#### 4) Web Scraping

Según Murillo y Saavedra en (2017) publicaron su artículo en el 4to Congreso Internacional AmITIC 2017, Aplicando nuevas tecnologías, en cual nos indican en que consiste la web scraping a continuación

El Web Scraping es una forma de minería de datos no estructurada, que permite extraer información de páginas web, escanear su código HTML y generar patrones de extracción de datos.

El artículo muestra las pruebas realizadas de estos métodos para medir la velocidad de extracción de los datos y buscar la mejor forma de extraer los datos de GS de forma estructurada.

Realizamos las pruebas utilizando 4 métodos: copiar y pegar, Local Browser, Local Software, Online. Ver Tabla 4.

Ha excepción del método de copiar y pegar, fue posible exportar los datos en formato .CSV, no sin antes realizar un proceso de depuración de los datos debido a que los datos que se extraen están unidos a otros textos que no eran de interés.

El artículo también muestra el análisis y desarrollo de un algoritmo en el lenguaje R, para comparar la velocidad de extracción de los datos y la eficiencia en el formato de salida de los datos.

Tabla 4:  
Método de web scraping, aplicación a utilizar y facilidad.

Métodos	Aplicación	Descarga	Conocimientos del usuario	Facilidad de uso
<b>Copiar y Pegar</b>	Manual	ninguno	ninguno	Fácil
<b>Web scraping Local Browser</b>	Extensión Chrome	Gratuito	Técnico mínimo	Fácil
<b>Web scraping Local Software</b>	Fminer	Pago (Trial)	Técnico intermedio	No es Fácil
<b>Web Scraping Online</b>	Import.io	Pago (Free versión)	Técnico mínimo	Fácil

Fuente 4to Congreso Internacional AmITIC 2017, Popayán  
Elaboración: Propia

### ✓ Aplicaciones utilizadas para cada método

#### 1. Copiar y Pegar:

No es un método de Web Scraping, la forma que todos alguna vez hemos realizado para obtener información de un sitio web, el proceso consistió en copiar y pegar cada dato del perfil y las publicaciones en una tabla de Excel, seleccionando solo el dato que se necesitaban, pero el trabajo resulto muy extenso.

#### 2. Web scraping Local:

Este método debe ser ejecutado por un usuario de forma manual utilizando la extensión SCRAPER de Google Chrome permite seleccionar un bloque de datos de una página web y al activar la extensión, extrae los datos que tengan el mismo

patrón de la clase HTML seleccionada, tener en cuenta se permite scrapear los datos una página por vez del perfil de Afiliación.

### **3. Web scraping Local Software:**

Este método es semiautomático debido que una persona debe indicar los datos que desea obtener y guardar a través del software FMiner, al abrir la página web en la aplicación, permite grabar el proceso como un macro donde se va creando un diagrama de flujo de datos de la página web asignando el valor seleccionado a cada variable.

### **4. Web scraping Online:**

Import.io es una aplicación Online que analiza automáticamente la estructura de la página web y muestra los datos en formato de tabla, es posible extraer datos de paginación, sin embargo, por lo que aumentaba el tiempo de extracción de los datos.(Murillo & Saavedra, 2017)

### **5) Hosting**

(Lopez B. 2018), quien indica en su artículo publica sobre que es un hosting lo siguiente.

Un hosting web es un servicio donde un proveedor te alquila un servidor conectado a Internet.

El uso más típico de un hosting es crear un sitio web (que, en realidad, no es más que un conjunto de ficheros en formato HTML que son las páginas web), pero también puedes usar tu hosting simplemente para permitir la descarga de cualquier otra cosa (documentos PDF, ficheros MP3 de audio, vídeo, etc.).

Una cosa muy importante a que, aparte de los servicios básicos de alojamiento de fichero, un servicio de hosting incluye otros servicios de mucho valor añadido. (Ver Tabla 5).

Existe una gama bastante extensa de tipos de alojamiento como puedes comprobar fácilmente en webs de comparativas como, por ejemplo, Hosting Mundo.

Tabla 5  
Servicios de Hosting.

---

<b>Servicios de mucho valor añadido del HOSTING</b>	
<b>Servidor de Correo</b>	Permite que tengas cuentas de correo con tu propio nombre de dominio.
<b>Aplicaciones web</b>	PHP y BDs para crear webs generalistas, blogs, tiendas online o foros de discusión, por citar algunas de las apps más importantes.
<b>FTP</b>	Para almacenar y descargar ficheros.
<b>Discos virtuales</b>	Crear almacenamiento en cloud con tu propio servicio de hosting al que accedes como si lo tuvieras en tu ordenador.

---

Fuente: Lopez,B  
Elaboración: Propia

Las principales opciones para la enorme mayoría de los usuarios, serán las de la lista siguiente. (Ver tabla 6)



Tabla 6  
Tipos de Hosting.

Tipos de Hosting		
Hosting	Ventajas	Desventajas
<b>Gratuito</b>	<ul style="list-style-type: none"> <li>▪ Para probar y aprender.</li> <li>▪ Permite mantener un entorno de desarrollo a coste cero.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Limitaciones de los recursos.</li> <li>▪ Nivel de servicio de muy baja calidad.</li> </ul>
<b>Compartido</b>	<ul style="list-style-type: none"> <li>▪ Necesidades de uso personal.</li> <li>▪ Resulta muy económica, soporta mucho tráfico.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Inestabilidad (caídas frecuentes y largas).</li> <li>▪ Mucho cuidado con el proveedor</li> </ul>
<b>Vps</b>	<ul style="list-style-type: none"> <li>▪ Es un esquema muy flexible.</li> <li>▪ No se puede ver comprometido por picos de actividad de otros clientes</li> </ul>	<ul style="list-style-type: none"> <li>▪ Aumento sensible del precio.</li> <li>▪ Mayor complejidad técnica de administración del servidor.</li> </ul>
<b>Especializado gestionado</b>	<ul style="list-style-type: none"> <li>▪ Servicios se encuentran optimizados.</li> <li>▪ Menos trabajo para el dueño del sitio web</li> </ul>	<ul style="list-style-type: none"> <li>▪ Aumenta el precio sensiblemente.</li> </ul>
<b>Cloud</b>	<ul style="list-style-type: none"> <li>▪ Seguridad.</li> <li>▪ Flexibilidad para crecer</li> </ul>	<ul style="list-style-type: none"> <li>▪ Precio más elevado.</li> </ul>

Fuente: Lopez,B

Elaboración: Propia

A la hora de elegir la opción que más de conviene hay que tener en cuenta, sobre todo, tres parámetros:

- Si la web se compone de simples páginas estáticas o se trata de una aplicación web como un blog WordPress.
- El tráfico que prevés para tu sitio web.
- Tus posibilidades económicas (personal o profesional).

(López B, 2018).

## 6) Aplicación Web

Las aplicaciones web actualmente son más populares ya que se utiliza en diferentes ámbitos y esto se debe a sus diferentes ventajas que un usuario puede tener.

Hoy en día los sistema operativos permiten ejecutar diversas aplicaciones multiplataformas y en cualquier dispositivo, las copias de seguridad son almacenadas en los servidores, la información que se genera puede ser

compartida de forma simultánea por varias personas, el espacio ocupado por los datos está a cargo del servidor y es de fácil uso. (Maritn & Martin, 2014)

### ✓ **Metodologías para el desarrollo de aplicaciones web.**

La metodología está compuesta por fases y estas pueden cambiar en el proceso que se desarrolla para evitar el menor número de errores y esto de una calidad y buen desempeño.

Las metodologías para aplicaciones Web contienen fases para el desarrollo de software que pueden aumentar o disminuir dependiendo del método que utilicen, según (Valle Rodriguez, 2009, pág. 17) la mayoría de los métodos coinciden en las siguientes etapas:

- Diseño Conceptual
- Diseño Navegacional
- Diseño de la presentación o diseño de interfaz de usuario

### ✓ **Fases De La Metodología Wsdm**

Los autores (Appelmans, 2004), (Aular & R. Talavera Pereira, 2007), (Chicaíza & Paredes, 2015), coinciden que la metodología WSDM, es una propuesta de desarrollo de aplicaciones Web que se enfoca en el usuario, y considera que este es el principal actor del sistema, es el encargado de definir los requisitos, y los usuarios que intervienen en la aplicación.

En WSDM se clasifica a los usuarios, y en base a ello se definen parámetros de diseño. Esta metodología se enfoca más en sitios Web estáticos que brindan información que sitios o aplicaciones dinámicas. Asimismo, no ha tenido tanta acogida, debido a que como se explicó anteriormente solo se centra en los usuarios y no tanto en los datos, otro aspecto importante en el desarrollo de aplicaciones Web.

La WSDM establece 4 fases o actividades que se especifican en la figura 6.

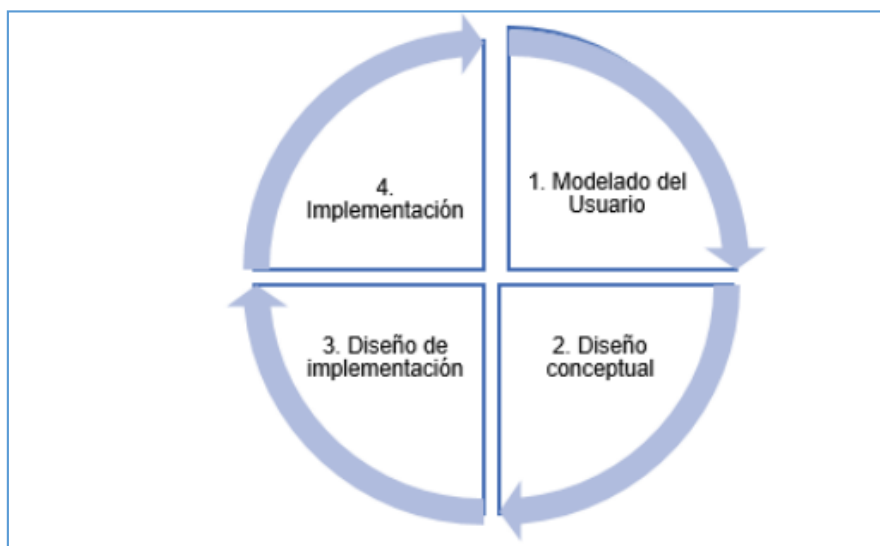


Figura 6: Fases de la Metodología WSDM.

Fuente: 3C Tecnología (Edición 25) Vol.7 – N° 1

A continuación, se realizará una breve descripción de las fases de esta metodología.

- ✓ **Modelado de Usuario:** sirve para identificar a los posibles usuarios de la aplicación y la información que ellos requerirían de este sitio.
- ✓ **Diseño conceptual:** se desarrolla el modelado conceptual, organiza la información, se clasifica a los usuarios, se modela los objetos, se crea diagramas entidad-relación y crea el diseño navegacional. Cada diseño de navegación en el sitio Web será diferente por cada perfil usuario y por ende tendrá su propia perspectiva. Los entregables de esta fase son el modelo conceptual, y diseño navegacional.
- ✓ **Diseño de implementación:** se crea un diseño en base a los requerimientos del usuario, este prototipo de interfaz del sitio Web deberá tener una apariencia agradable, ser eficiente y seguro, así mismo aquí se especifican las restricciones de diseño, según lo que se estableció en el diseño conceptual.
- ✓ **Implementación:** se realiza la selección del entorno de desarrollo, construcción de la arquitectura, codificación y verificación de la funcionalidad total de la aplicación Web.

✓ **Sohdm (Scenario-Based Object-Oriented Hypermedia Design Methodology)**

El Sohdm metodología orientada a objetos que se basa en diseños con distintos escenarios o panoramas.

A lo que Cobo agrega que permite capturar las necesidades del sistema proponiendo el uso de escenario. SOHDM parte de un diagrama donde se identifican las entidades externas capaces de comunicarse con el sistema, es una metodología muy parecida a la metodología OOHDM diferenciadas por la utilización de escenarios. (Cobo, 2011)

En cuanto a los procesos de gestión de desarrollo de software o ciclo de vida se describen 6 fases las cuales se pueden observar en la figura 7.

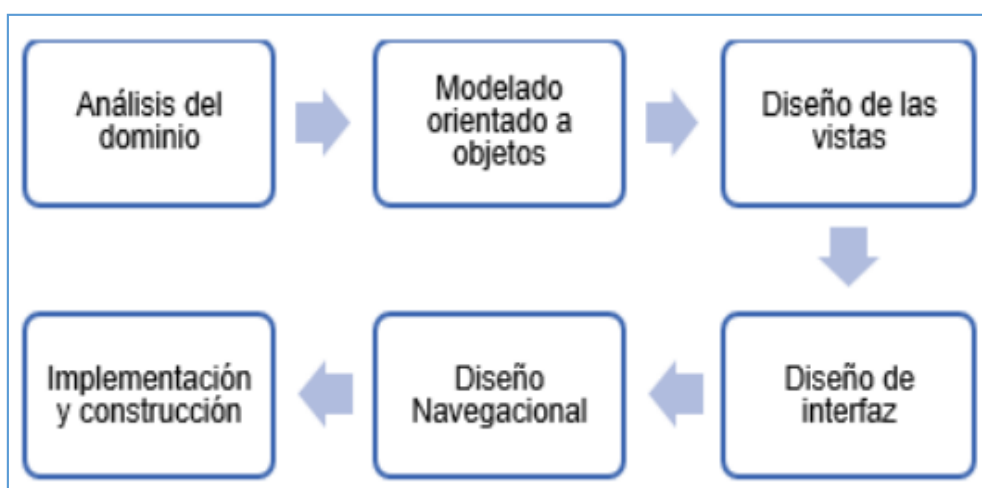


Figura 7: Fases de la metodología SOHDM

Fuente: & Elaboración: 3C Tecnología (Edición 25) Vol.7 – N° 1

- ✓ Análisis del dominio: establece los límites de la aplicación que se desarrollará, y se los representa mediante un diagrama de flujo. Además, se hace uso de los SACs (Scenarios activity charts) que no son más que escenarios donde se determina los requisitos de la aplicación.
- ✓ Modelo de objetos: en esta etapa se utilizan los SACs para realizar el modelado de objetos. Los usuarios son los principales objetos del

sistema, cada usuario es descrito en el documento de desarrollo de la aplicación, los cuales incluyen atributos, asociaciones y cardinalidad.

✓ **Oohdm (Object Oriented Hypermedia Design Methodology)**

Esta metodología permite desarrollar aplicaciones Web a partir de la utilización de modelos especializados como: conceptual, navegación e interfaz de usuario teniendo como objetivo simplificar y hacer más eficaz el diseño de aplicaciones.

A continuación, se describirán las cinco etapas de la metodología OOHDM. (Ver la figura 8)

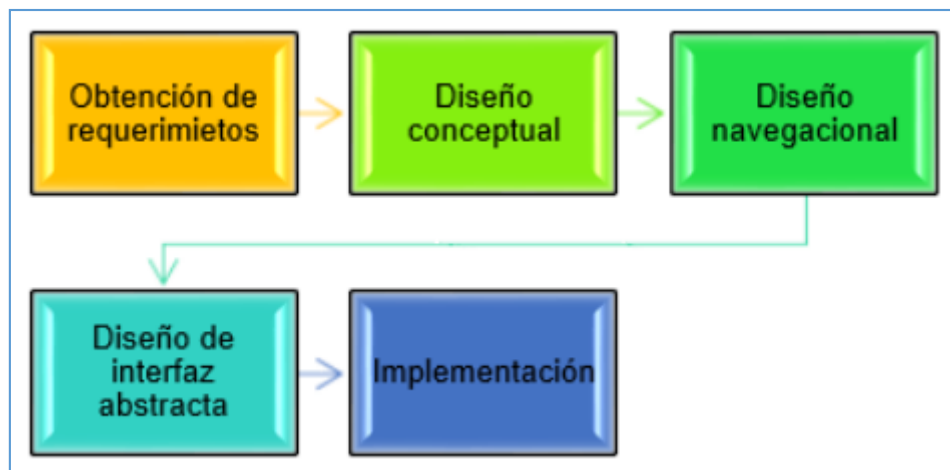


Figura 8: Fases de la metodología OOHDM

Fuente: & Elaboración: 3C Tecnología (Edición 25) Vol.7 – N° 1

Es importante mencionar que el desarrollador y diseñador son los encargados del lado técnico del sistema y su apariencia final, mientras que el cliente verifica que funcione correctamente como lo ha solicitado en la primera fase o etapa.

✓ **Wae (web application extension).**

La WAE es una extensión de UML, que no se enfoca en el paradigma orientado a objetos si no en los elementos Web. WAE incorpora algunos conceptos como JavaScript y Form.

En esta metodología cubre el lado tanto del servidor (páginas del servidor) como el cliente (Active X, applet Java, etc). Sin embargo, los conceptos

orientados a objetos (por ejemplo, herencia) no están suficientemente preocupados por la extensión.

Según el análisis de (Bastidas, 2009), la metodología WAE cuenta con las siguientes fases como se indica en la figura 9.

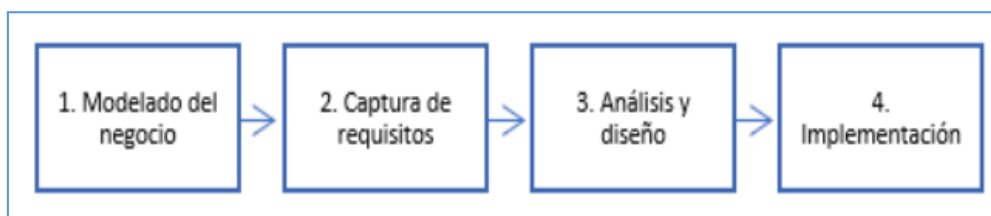


Figura 9: Fases de la metodología WAE

Fuente: & Elaboración: 3C Tecnología (Edición 25) Vol.7 – N° 1

#### ✓ Iweb (ingeniería web).

“La IWeb se divide en un número de actividades estructurales, también llamadas regiones de tareas. Generalmente, existen entre tres y seis regiones de tareas, las cuales no necesariamente se deben aplicar todas por cada iteración” (Bermúdez, 2008). IWeb es una metodología que se enfoca en la creación de aplicación y sistemas Web de alta calidad, basándose en principios científicos de ingeniería. Dichas aplicaciones hacen posible el acceso desde ordenadores remotos.

A continuación, citaremos las siete fases de la metodología IWeb. (Ver Figura 10)

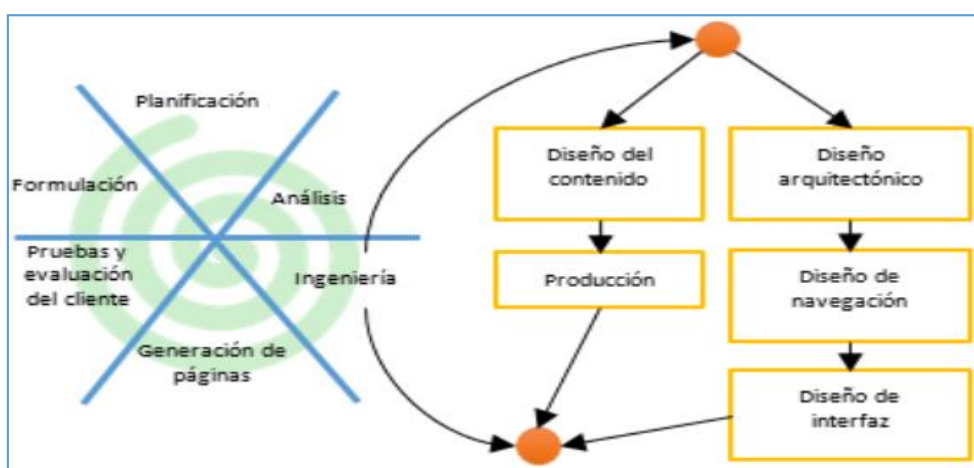


Figura 10: Fases de la metodología IWEB

Fuente: & Elaboración: 3C Tecnología (Edición 25) Vol.7 – N° 1

- ✓ **Formulación:** En esta primera fase se identifican los objetivos, metas, se establece el alcance de la aplicación y su primera entrega. Aspectos muy importantes a considerar en esta fase es conocer la motivación para el desarrollo de la aplicación, verificar si es necesaria o no, e identificar quien la va a utilizar.
- ✓ **Planificación:** estimar el coste general del proyecto a realizar, así como también planes de contingencia debido a posibles riesgos, el ámbito y describir la calidad y gestión de la aplicación en cuanto a cambios.
- ✓ **Análisis:** establecer los requerimientos de diseño y técnicos, también se analiza el contenido del mismo, su iteración, funcionalidad y configuración.
- ✓ **Ingeniería:** lograr la integración del diseño arquitectónico, de navegación y de interfaz. Es por esto que se divide en diseño del contenido, producción, diseño arquitectónico diseño de navegación, y diseño de interfaz.
  
- ✓ **Generación de páginas:** se integran los diseños de la etapa anterior a través de herramientas como lenguajes de programación y etiquetado que sirvan como base la construcción de la aplicación Web.
- ✓ **Pruebas o test:** se comprueba la lógica de negocios aplicada en el sistema, y se verifican las entradas y salidas de datos con el fin de descubrir errores de funcionalidad, comportamiento o rendimiento.
- ✓ **Evaluación de cliente:** permite corregir errores gracias a las iteraciones realizadas con el fin de ir puliendo la aplicación en comparación a las iteraciones anteriores.

Observar a continuación, diferentes criterios de comparativas de metodologías donde se abarcan diferentes criterios (Ver figura 11)

Requerimientos	Metodologías					
	WSDM	SOHDM	OODHM	UWE	WAE	IWEB
Datos	X	X	X	X	X	X
Interfaz de usuario		X	X	X	X	X
Navegacionales			X	X	X	X
Personalización	X		X			
Transaccionales		X		X		
No funcionales	X	X	X	X	X	X

Figura 11: Comparación de requisitos en el entorno Web en las metodologías.

Fuente: & Elaboración: 3C Tecnología (Edición 25) Vol.7 – N° 1

Analizando los resultados y teniendo en cuenta que las metodologías, se encuentran ordenadas en orden cronológico. Se puede evidenciar que en un principio las metodologías solo se centraban en los datos y la interfaz que se le proporcionaba al usuario mientras que en las más actuales se resalta la importancia de tratar los requisitos de personalización, navegación, transaccionales y no funcionales. (Molina Ríos, 2018)

## 7) Proceso de Publicar

(Brinder, 2016, págs. 1-10), en su artículo nos explica de cómo se hace el paso a paso del proceso de publicar sede una noticia hasta un libro y lo hace de la siguiente manera:

Los materiales impresos tienen otra complejidad (¡y costo!). Por eso nos enfocaremos en publicaciones digitales que podremos diseñar con tan sólo una computadora y software libre accesible para todos y todas.

- Informes, boletines, libros digitales en PDF, libros electrónicos
- resúmenes de actividades, fanzines, folletos, revistas, etc.

### ✓ ¿Cómo es el paso a paso?

#### a) Redacción

Esta primera etapa es la de la creación del texto puro y duro. ¿Qué es lo que queremos comunicar en nuestro documento? Contestándonos esta pregunta sabremos la estructura que deberá tener.



Podemos hacer un informe con los resultados de una investigación o contar a nuestra comunidad las actividades que nuestra radio u organización llevó a cabo a lo largo del año, por ejemplo. (Ver figura 12)

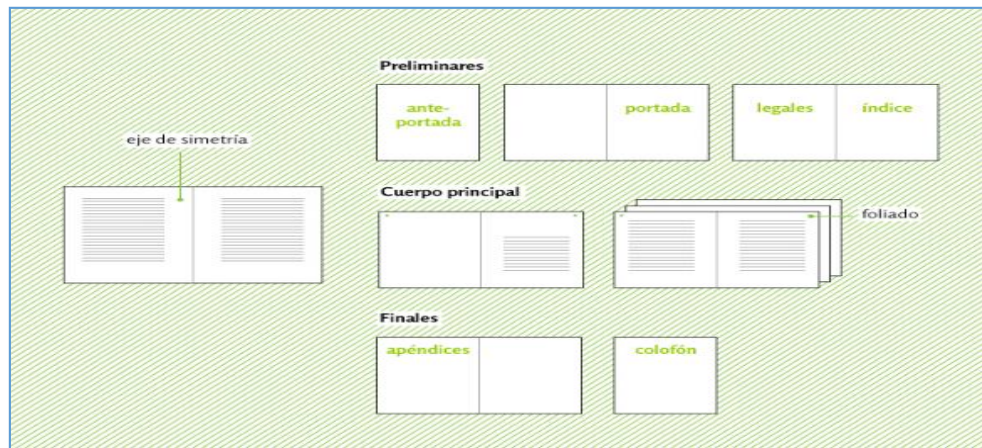


Figura 12: Estructura del libro

Fuente: & Elaboración: Open Educational Resources for Typography

En términos generales los documentos cuentan con los siguientes elementos:

- ✓ Tapa, es la cara de presentación de nuestro documento por lo que en ella deben estar contemplada una serie de datos básicos como el título, las y los autores, el nombre de la radio, organización o editorial que está publicando.
- ✓ Anteportada o portadilla, es la primera hoja del documento como tal. Se repiten el título, las y los autores, y la radio u organización que publica.
- ✓ Hoja de legales: aunque ya lo veremos más en detalle en el capítulo 3, aquí se detallan la catalogación, el ISBN, los datos de quien publica, el año de publicación, la licencia, etc.
- ✓ Índice: es una guía del detalle de la estructura del documento para facilitar su lectura.
- ✓ Prólogo: son unas palabras sobre el documento, el proceso de publicación, y quienes participaron de él.
- ✓ Introducción: en esta sección hace un recorrido por los contenidos del documento y se adelantan las primeras conclusiones.
- ✓ Capítulos: son unidades temáticas del documento. Podemos armar cuantos capítulos necesitemos.

- ✓ Referencias bibliográficas: pueden estar en cada uno de los capítulos, como una nota al pie, o podemos recopilarlas todas al final del documento. Existen distintas normas para escribir las referencias. La más difundida es APA, la Asociación Americana de Psicología. Para saber cómo utilizar este sistema puedes hacer clic aquí.
  
- ✓ Contraportada: ¿quién no ha ido por ahí de compras a una librería de usados y lo primero que hace luego de ver la portada es dar vuelta el libro para enterarnos un poco más de qué se trata? Pues para eso sirve la contraportada. Aunque pierde un poco el sentido cuando se trata, como en este caso, de documentos digitales, debemos mantenerla ya que es el cierre de la publicación.

Lo ideal es que haya una persona responsable de esta iniciativa para que el texto tenga coherencia y un mismo registro.

#### **b) Edición.**

El proceso de edición es complejo y no hay que subestimarlo. Desde que tenemos la versión final del archivo de texto hasta su publicación hay un largo trecho. Es importante que cuando planifiquemos el proceso de publicación le asignemos el tiempo que necesita. Acortar plazos puede redundar en una publicación de menor calidad.

La edición no se trata únicamente de revisar que no haya errores de ortografía. Se trata de una visión global sobre el documento, como un escultor que va dando forma a un bloque de mármol.

### c) Corrección

- ✓ Consulta el diccionario siempre que tengas una duda sobre cómo se escribe una palabra.
- ✓ Una ortografía impecable es parte importante de la calidad de una publicación.
- ✓ En títulos y subtítulos sólo la primera palabra lleva mayúscula (salvo que incluya nombres propios). Además, no llevan punto final.
- ✓ El sujeto y el predicado no se separan por coma.
- ✓ Se utiliza mayúsculas para la primera palabra de la oración luego de un punto, nombres propios, cargos (cuando no están acompañados del nombre de quien lo ostenta).
- ✓ Las comillas se utilizan para las citas o textos irónicos, pero nunca para resaltar o dar más importancia a una oración.
- ✓ Palabras en otros idiomas se escriben en cursiva.

### d) Diseño y diagramación.

El diseño de las tapas y la diagramación de los interiores transformarán nuestro archivo de procesador de textos en una publicación con todas las letras. Hay programas privativos superpotentes que si bien ofrecen todas las herramientas que necesitamos tienen licencias de uso que difícilmente puedan asumir nuestras radios y organizaciones.

### e) Impresión

- ✓ Existen muchos métodos de impresión. Aquí no los veremos en detalle porque trabajaremos exclusivamente sobre publicaciones digitales (PDF y ePUB). Pero las más difundidas son:
- ✓ Artística: son aquellos donde la imagen se graba manualmente y luego se hacen reproducciones limitadas de manera manual.
- ✓ Digital: es aquel en el que la impresión se realiza directamente al papel desde un documento digital. Puede ser láser o a inyección.
- ✓ Industrial: son aquellos que necesitan planchas, placas o mallas de algún tipo, para realizar la impresión. Las más conocidas son offset o la serigrafía.

## **f) Distribución**

Aunque parezca evidente que luego de publicar nuestros documentos tengamos que distribuirlos, esta no es una etapa menor. Muchas veces terminamos buscando depósitos para guardar las cajas de libros nuevitos, o subimos nuestro PDF al sitio web de la radio u organización y nadie lo lee.

Debemos pensar estrategias para que nuestra publicación llegue a quienes está dirigido y así todo nuestro trabajo no haya sido en vano. Por ejemplo, ¿qué te parece organizar una presentación o armar una mesa en tu organización? También se pueden sortear ejemplares entre la audiencia. (Brinder, 2016, págs. 1-10)

### **1.9 Definición de términos básicos.**

#### **1) Aplicación Web. -**

Según Rafael Mendez una aplicación web es generalmente una manera de facilitar el logro de una tarea específica en la Web, a diferencia de un sitio web estático que es más bien una herramienta, no menos importante, para la comunicación. El término más decisivo de esta definición es "tarea específica". La aplicación web por lo tanto permite al usuario interactuar directamente contigo y tus datos, todo en forma personalizada ("Rafael Menéndez-Barzanallana Asensio. Historia del desarrollo de aplicaciones Web. Universidad de Murcia," n.d.).

#### **2) Publicaciones periódicas. -**

Diarios, revistas, fascículos, etc. son registros parciales de nuestra cultura con cierto grado de relatividad temporal y espacial. Estas publicaciones reproducen contenido e informaciones renovables: se actualizan en el período de tiempo que corresponde (diario, semanal, quincenal, mensual, etc.) por tanto, deben contar a la hora de su lanzamiento con una maqueta inicial (Pablo Cosgaya, Natalia Pano, 2012)

**3) Medio de comunicación. -**

Tiene como objetivo informar acerca de los acontecimientos de relevancia y trascendencia, con noticias del ámbito local, nacional e internacional; además, orienta y determina ideas, actitudes y comportamientos individuales, grupales, y de clase social. (Goya, n.d.)

**4) Web Scraping.-**

La recopilación automatizada de datos de Internet es casi tan antigua como la propia Internet.

Aunque el raspado web no es un término nuevo, en años pasados la práctica ha sido más comúnmente conocido como raspado de pantalla, minería de datos, recolección de datos en la web o variaciones similares. En la práctica, el web scraping abarca una amplia variedad de técnicas de programación y Tecnologías, como el análisis de datos y la seguridad de la información. (Mitchell, 2015).

**5) Desarrollo de sistemas. -**

El desarrollo de sistemas es un proceso para la resolución de problemas consistente en investigar una situación, diseñar una solución que mejore la situación, adquirir los recursos humanos, financieros y tecnológicos necesarios y, finalmente, evaluar el éxito de la solución. (BEEKMAN, 2005).

**6) Directorio. -**

Un directorio web es un directorio en el World Wide Web que publica enlaces a otros sitios web y clasifica esos enlaces por diversos temas, por lo que los usuarios de Internet sin experiencia pueden encontrarlos fácilmente. (Chung, 2012).

**7) Web. -**

La World Wide Web, más conocida como Web, es una de las áreas de internet que se ha desarrollado más rápidamente. Nació en 1989, como parte de un proyecto del CERN de Suiza y con el objetivo de mejorar el

intercambio de información dentro de internet, y vea en que se ha convertido actualmente. (Hobbs, 1999).

#### **8) Sistema de Información. -**

Conjunto formal de procesos que, operando sobre una colección de datos estructurada de acuerdo a las necesidades de la empresa, recopila, elabora y distribuyen selectivamente la información necesaria para la operación de dicha empresa y para las actividades de dirección y control correspondientes, apoyando, al menos en parte, los procesos de toma de decisiones necesarios para desempeñar funciones de negocio de la empresa de acuerdo con su estrategia. (ANDREU, E., & VALOR, 1991).

#### **9) Extracción de información. -**

La extracción de información es un proceso que a partir de textos logra extraer automáticamente datos ordenados y poco ambiguos, respecto a un campo en específico, lo cual incluye entidades, relaciones y eventos de los mismos. (Abolhassani, 2003).

#### **10) Html.-**

Son las siglas de Hyper Text Markup Language, o lenguaje de marcas de hipertexto. Es un estándar que define una estructura básica y un código (denominado código HTML) para la definición de contenido de una página web, como texto, imágenes, etc. Este lenguaje no describe la apariencia ni el funcionamiento del sitio. (Mora, 2002).

#### **11) JavaScript.-**

Es un lenguaje interpretado usado para múltiples propósitos, pero solo considerado como un complemento hasta ahora. Una de las innovaciones que ayudó a cambiar el modo en que vemos Javascript fue el desarrollo de nuevos motores de interpretación, creados para acelerar el procesamiento de código. La clave de los motores más exitosos fue transformar el código Javascript en código máquina para lograr velocidades de ejecución similares a aquellas encontradas en aplicaciones de escritorio. Esta mejorada capacidad permitió superar viejas

limitaciones de rendimiento y confirmar el lenguaje Javascript como la mejor opción para la web. (Gauchat, 2012).

#### **12) RSS.-**

Es un formato de datos que sirve para el envío de contenidos a quienes están registrados en un determinado sitio de Internet. Esta estructura permite que la distribución del contenido se realice sin que sea necesario valerse de un navegador, ya que la acción se lleva a cabo a través de un software creado especialmente para leer esta clase de datos que se conoce como agregador. De todas formas, las versiones más recientes de los principales navegadores contemplan la posibilidad de leer los contenidos de RSS sin necesidad de acudir a un software adicional. (Gardey, 2008).

#### **13) MongoDB.-**

Es una base de datos orientada a documentos. Esto quiere decir que, en lugar de guardar los datos en registros, guarda los datos en documentos. Estos documentos son almacenados en BSON, que es una representación binaria de JSON. (Monrraga, 2014).

#### **14) Frontend.-**

Frontend es la parte de un programa o dispositivo a la que un usuario puede acceder directamente. Son todas las tecnologías de diseño y desarrollo web que corren en el navegador y que se encargan de la interactividad con los usuarios. (Chapaval, 2018).

#### **15) Backend.-**

Es la capa de acceso a datos de un software o cualquier dispositivo, que no es directamente accesible por los usuarios, además contiene la lógica de la aplicación que maneja dichos datos. El Backend también accede al servidor, que es una aplicación especializada que entiende la forma como el navegador solicita cosas. (Chapaval, 2018).

**16) URL.-**

Es una sigla del idioma inglés correspondiente a Uniform Resource Locator (Localizador Uniforme de Recursos), por lo tanto, es el conjunto de caracteres que posibilita la asignación de una dirección exclusiva a un recurso que se encuentra disponible en el espacio virtual. En otras palabras, el URL es una dirección de Internet que, al ser encontrada y visualizada por un navegador, muestra un recurso de información al usuario. (Merino, 2012).

**17) Tecnología.-**

La tecnología es un conjunto de nociones y conocimientos utilizados para lograr un objetivo preciso, la solución de un problema específico del individuo o a la satisfacción de alguna de sus necesidades. Es un concepto extremadamente amplio que es capaz de abarcar una inmensa variedad de aspectos que pueden ir de la electrónica al arte o la medicina. (Raffino, 2019).

**18) Calidad.-**

La definición de calidad más aceptada en la actualidad es la que compara las expectativas de los clientes con su percepción del servicio. El desarrollo de la industria de los servicios ha supuesto un desarrollo de una nueva óptica del concepto de calidad que se focaliza más hacia la visión del cliente (García, 2001).

**19) Usabilidad.-**

La usabilidad se refiere a la capacidad de un software de ser comprendido, aprendido, usado y ser atractivo para el usuario, en condiciones específicas de uso. (ISO/IEC 9126).

**20) Trafico web.-**

Según Sebastián P. se determina aquellas visitas de usuarios que llegan a la plataforma digital (web o blog profesional) del emprendimiento o proyecto digital en cuestión. (Sebastián P. 2018).



## **II. MÉTODO**

### **2.1 Tipo y Diseño de investigación**

#### **2.1.1 Diseño de la investigación**

En la investigación se aplicó el Diseño– Cuasi Experimental,

Donde se describe que es una investigación que posee todos los componentes de un experimento, excepto que los sujetos no se asignan aleatoriamente a los grupos. En carencia de aleatorización, el investigador se enfrenta con la tarea de identificar y separar las consecuencias de los tratamientos del resto de factores que afectan a la variable dependiente. (Pedhazur & Schmelkin, 1991).

#### **2.1.2 Tipo de investigación**

El tipo de investigación que se utilizó en la tesis es aplicada por que la variable independiente (Aplicación web Scraping) es la que se va implementar y de esta manera se podrá demostrar lo que sucederá en las variables dependientes.

Esto lleva la necesidad de repensar la coordinación en la formación profesional, para unificar criterios y establecer una profunda unión e integración con base en el diálogo, para poner en común “verdades” distintas sobre la realidad que se investiga, que permita, a su vez, descubrir significados a partir de la discusión dentro de comunidades científicas. (Martinez, 2004, págs. 59-72)

#### **2.1.3 Nivel de la investigación**

En el desarrollo de la tesis el nivel de la investigación es explicativa ya que la causal es la variable dependiente (publicación de noticias tecnológicas en el portal web ojotecnologico) está direccionado a la variable independiente (Aplicación web scraping).

Hay predominio de explicación, descripción y correlación. Es aquella que tiene relación causal, no solo persigue describir o acercarse a un problema, sino que intenta encontrar las causas del mismo. En este caso la dirección es V.D. → V.I. (Sánchez & Reyes, 2006)

#### **2.1.4 Enfoque de la investigación**

El desarrollo de la tesis tiene un enfoque cuantitativo, porque tiene una secuencia de pasos consecutivos y lógicos.

Según Hernández F. las preguntas se establecen hipótesis y determinan variables; se desarrolla un plan para probarlas (diseño); se miden las variables en un determinado contexto; se argumentan las mediciones obtenidas (con frecuencia utilizando métodos estadísticos), y se establece una serie de conclusiones respecto de la(s) hipótesis. (Hernández, Fernández, & Baptista, 2010, págs. 7,Cap.1)

## **2.2 Población, Muestra y muestreo**

### **2.2.1 Población**

Como mencionan, para llevar a cabo un buen muestreo de la investigación primero hay que definir la población o Universo, que es el conjunto de objetos, hechos o eventos que se van a estudiar con las variadas técnicas que hemos analizado. Siguiendo con su explicación, la muestra es el subconjunto o parte de la población o Universo que habiendo sido seleccionado por diversos métodos y teniendo en cuenta la representatividad del universo, será representativa si es que reúne las características que poseen los individuos del Universo o Población. (Ñaupas, Mejia, Novoa, & Villagomez, 2014, pág. 246).

El total de nuestra población es lo 2000 seguidores que tenemos actualmente en nuestra fanpage de Facebook el cual siguen nuestras noticias publicadas, y a su vez cada día va aumentando los seguidores.

Y para la investigación se seleccionó una población que está conformada por 80 seguidores en nuestro fanpage de Facebook, el cual han sido seleccionados por las siguientes condiciones: compartidos, likes o reacciones a las publicaciones, los que comentan frecuentemente, etiquetados entre otros.

### 2.2.2 Muestra

- Para el proceso cuantitativo, la muestra es un subgrupo de la población de interés sobre el cual se recolectarán datos, y que tiene que definirse y delimitarse de antemano con precisión.
- El investigador pretende que los resultados encontrados en la muestra se generalicen o extrapolen a la población. (Hernández, Fernández, & Baptista, 2010)

Muestra de nuestra investigación cuando se conoce el tamaño de la población.

Se utilizar la siguiente formula:

$$n = \frac{Z^2 PQN}{(N - 1)E^2 + Z^2PQ}$$

En donde:

n = Tamaño de la muestra.

Z = Valor Z curva normal (1.96)

P = Probabilidad de éxito del (0.5)

Q = Probabilidad de fracaso (0.5)

N = Población de (80)

E = Error muestral de (00.5)

### 2.2.3 Muestreo

**Sustituyendo la fórmula:**

$$n = \frac{1.96^2 * (0.5)(0.5) * 80}{(80 - 1) * (0.05)^2 + Z(1.96)^2 * (0.5)(0.5)}$$

$$n = \frac{76.83}{1.1579}$$

n = 66 (valor redondeado)

El total de la muestra de la investigación fue de 66 usuarios

## 2.3 Técnicas e instrumentos para la recolección de datos.

### 2.3.1 Técnicas

Las técnicas que se usaran en el trabajo de investigación de la tesis son los siguientes.

#### ✓ Encuestas

Mayntz et al., (1976:133) citados por Díaz de Rada (2001:13), describen a la encuesta como la búsqueda sistemática de información en la que el investigador pregunta a los investigados sobre los datos que desea obtener, y posteriormente reúne estos datos individuales para obtener durante la evaluación datos agregados.

Resumen de las diferentes técnicas de recolección de datos con el fin de obtener los resultados precisos y confiables para el contraste de hipótesis planteadas (Ver tabla 7)

Tabla 7  
Técnicas de recolección de datos.

<b>Indicador VD</b>	<b>Técnica</b>
<b>Número total de noticias por día</b>	
<b>Cantidad de vistas</b>	Encuesta
<b>Usabilidad</b>	

Fuente y elaboración propia

### 2.3.2 Instrumentos

Los instrumentos que se utilizara para la recolección de datos son:

- **Cuestionarios:** para la recolección de datos de la muestra se utilizó un cuestionario, que se elaboraron de acuerdo a la operacionalización de las variables que señala (Hernández-Sampieri y otros, 2014, p. 217), se abarcaron preguntas cerradas con el escalamiento de Likert en las opciones de respuesta, debido a las posibilidades de manejo durante la aplicación del instrumento, para el análisis de datos.

A continuación, se presenta el indicador con su respectivo instrumento con el fin de obtener los resultados precisos y confiables para el contraste de las hipótesis planteadas. (Ver Tabla 8)

Tabla 8  
Instrumento de recolección de datos.

<b>Indicador VD</b>	<b>Instrumento</b>
<b>Número total de noticias por día</b>	
<b>Cantidad de visitas por día</b>	Cuestionario
<b>Nivel de usabilidad</b>	

Fuente y elaboración propia

## 2.4 Valides y Confiabilidad de instrumentos

- ✓ **Confiabilidad de instrumentos**

El criterio que usara para la confiabilidad del instrumento es Método de consistencia interna (Cronbach).

Tabla 9  
Fiabilidad Alfa de Cronbach

<b>Estadísticas de fiabilidad</b>	
Alfa de Cronbach	N de elementos
,961	24

Fuente: Base de datos de la encuesta, analizado en SPSS v.25 2019

Tabla 10  
Fiabilidad Alfa de Cronbach total de los elementos

	<b>Estadísticas de total de elemento</b>			
	Media de escala si el elemento se ha suprimido	Varianza de escala si el elemento se ha suprimido	Correlación total de elementos corregida	Alfa de Cronbach si el elemento se ha suprimido
P1. ¿Según su criterio la cantidad de noticias publicadas por día es la correcta?	83,46	364,953	,651	,960
P2. ¿A su criterio las fuentes de comunicación que se muestran actualmente es suficiente o se debería aumentar más?	83,47	367,396	,636	,960
P3. ¿Es importante saber que fuente de comunicación pública más noticias por día?	83,50	364,557	,695	,960
P4. ¿Está de acuerdo que solo encuentra noticias únicas y no duplicadas?	83,40	369,082	,640	,960
P5. ¿A su opinión se ve un incremento de visitas al portal web?	84,14	345,178	,846	,958
P6. ¿A su criterio ha aumentado el número de usuarios recurrentes?	84,03	347,114	,829	,958
P7. ¿La automatización de la publicación de noticias ha tenido un impacto positivo en la afluencia de usuarios?	84,08	341,588	,860	,958
P8. ¿Generar informes personalizadas para llevar el control de vistas diarias?	84,09	343,183	,869	,958

P9. ¿La mayoría de personas aprendería a utilizar nuestra web de noticias tecnológicas?	83,46	367,228	,632	,960
P10. ¿Se siente seguro al navegar por nuestra web?	83,33	369,641	,656	,960
P11. ¿Cree usted que es bastante fácil navegar vía web y dispositivo móvil?	83,38	371,046	,602	,961
P12. ¿Cree que las diferentes funciones de nuestra web se encuentran muy bien integradas?	83,44	367,012	,675	,960
P13. ¿El aspecto de nuestra web es el adecuado para mostrar las noticias tecnológicas?	84,28	348,386	,794	,959
P14. ¿La página web transmite la idea del proyecto?	84,35	351,740	,775	,959
P15. ¿El portal ojotecnologico es una web recomendable?	84,45	353,928	,744	,959
P16. ¿Según tu criterio la información que se muestra en Ojotecnologico es relevante?	84,24	354,139	,730	,960
P17. ¿Resulta más rápido y seguro la automatización de publicación de noticias?	83,48	366,068	,647	,960
P18. ¿La aplicación web scraping, ha hecho que se reduzcan los costos de publicación?	83,46	367,090	,605	,961
P19. ¿La tecnología genera progreso para ojotecnologico?	83,47	365,549	,606	,961
P20. ¿Cree usted que se reducirá personal pro la automatización de publicación de noticias tecnológicas?	83,43	363,866	,653	,960
P21. ¿El Buscador de noticias localiza fácilmente la información?	83,47	365,304	,708	,960
P22. ¿La actualización de noticias en nuestra web es constante?	83,53	362,098	,662	,960
P23. ¿Desde su punto de vista existe facilidad de acceso a la información en el portal ojotecnologico?	83,35	370,107	,631	,960
P24. ¿El tiempo de carga de la web es lo que usted esperaba?	83,54	362,800	,656	,960

Fuente: BD de la encuesta, analizado en el software SPSS v.25 2019

### ✓ Validez de instrumentos

Se usarán los siguientes criterios para la valides del instrumento.

- Valides de contenido.
- Valides de criterio.

## 2.5 Método de análisis de datos

Se inicio con el análisis de datos en cuento se terminó de realizar las encuestas sobre la web scraping para la mejora de publicacion de noticias en el portal de noticias ojotecnologico, se utilizó el software estadístico para ciencias sociales (Statistical Package for Social Science – SPSS, versión 25), con el cual se realizaron los análisis estadísticos descriptivos (tablas de frecuencias, gráficas, medidas de variación) e inferenciales (análisis correlacional y prueba de hipótesis).

A continuación, se muestra las técnicas de procesamiento y el análisis de la información (ver Tabla 11)

Tabla 11  
Matriz análisis de Datos.

Variable	Indicador	Escala de medición	Estadísticos descriptivos	Análisis inferencial
Aplicación web scraping	Cantidad de noticias publicadas por día	Escala de Proporción / Razón	Tendencia central (media aritmética, mediana y moda) Dispersión (varianza, desviación estándar) Posición (cuartiles) Forma (asimetría y curtosis)	Prueba paramétrica (T student)
	Cantidad de vistas por día	Escala Ordinal	Tendencia central (media aritmética, mediana y moda) Dispersión (varianza, desviación estándar) Posición (cuartiles) Forma (asimetría y curtosis)	Prueba paramétrica (T student)



Nivel de usabilidad	Escala de Proporción / Razón	Tendencia central (media aritmética, mediana y moda) Dispersión (varianza, desviación estándar) Posición (cuartiles) Forma (asimetría y curtosis)	Prueba paramétrica (T student)
---------------------	------------------------------	--	--------------------------------

Fuente y elaboración propia

## 2.6 Aspectos Éticos

Para sustentar que la investigación cuente con los principios de la ética, porque los encuestados son personas, se tendrá como premisa el consentimiento previo de los mismos para participar, tomándose en cuenta la protección de datos personales con los aspectos establecidos al respecto.

Se tomará en cuenta si las políticas públicas hacen posible el desarrollo de la investigación, si es factible estudiar el planteamiento en cuestión, para ello se cuenta con los recursos necesarios para la misma, yo como investigador soy competente para realizar ese tipo de estudio, con lo cual veo pertinente y con el consentimiento informare a las personas involucradas en la investigación.

### III. RESULTADOS

#### 3.1 Resultados descriptivos

##### 3.1.1 Descripción del proceso manual de publicación

El objetivo de este punto es exponer la técnica de scraping para poder automatizar el proceso de recolección, filtrado, validación y publicación de noticias el cual permitirá brindar mayor satisfacción al usuario lector y reducir los costos que acarrea la publicación de noticias de forma manual.

Adicional se realizará un análisis de la situación actual de cada hipótesis y luego aplicando la teoría se mostrarán los resultados obtenidos.

Análisis de la situación actual de cómo el proceso de publicación de noticias es de manera manual y tomó tiempo el cual no permite que se agreguen un gran número de noticias, a continuación, se detalla el proceso de publicación. (Ver Imagen 13)

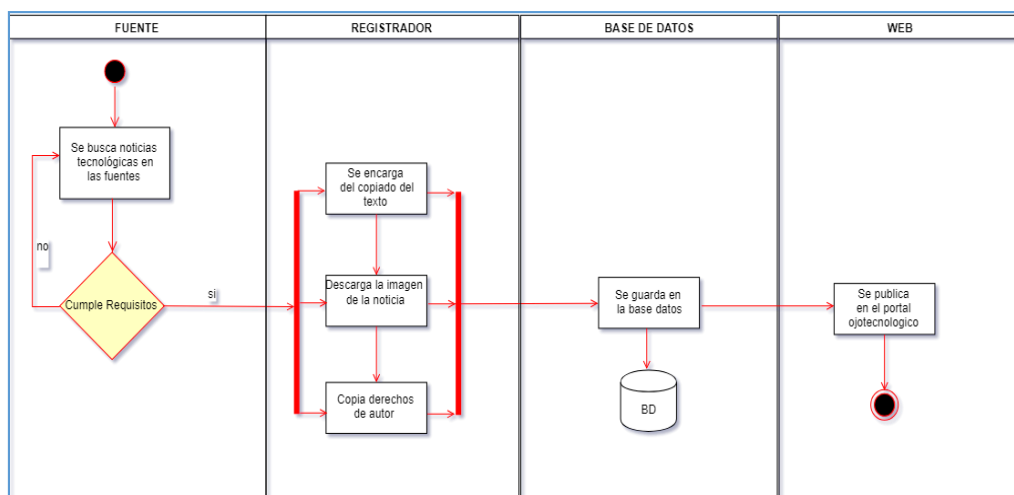


Figura 13: Diagrama de flujo del proceso de publicación de noticias

Fuente: Ojotecnologico

Elaboración: propia

Para publicar una noticia en el portal ojotecnologico, el primer paso que tenemos que realizar es la búsqueda de noticias tecnológicas y con información actualizada, ingresamos a los medios de comunicación que ya tenemos registrados,

Las noticias tecnológicas son revisadas verificando que cuenten como mínimo con una imagen para ser publicadas en ojotecnologico, el inconveniente que tenemos es que debemos revisar noticias por noticia y hacer un filtro manual para ser descartada o si la noticia es muy relevante en cuanto a tecnología se le busca una imagen de internet que vaya acorde al tema de la noticia y de esta manera recién ser guardada en la base de datos para su publicación.

El copiado de noticias lo hacemos de forma manual somos tres personas las que le dedicamos tiempo de acuerdo a nuestra disponibilidad, y esto ocasiona que obviemos texto al momento del copiado de la noticia y nos damos cuenta ya cuando las noticias fueron publicadas.

El tiempo que empleamos para guardar las noticias en la base de datos es alto debido a la cantidad de noticias por día que se pueda insertar, si eran pocas lo hacíamos manualmente, y en otros casos utilizamos un Excel para hacer una importación directa a la BD, el cual en ocasiones presentamos problemas o errores que nos retrasa en la publicación de noticias.

En cuanto a la cantidad de noticias publicadas por día si tenemos un serio problema debido que, al tener pocas fuentes de información, ya que el trabajo de publicar lo realizamos de forma manual y somos tres personas que trabajamos a diario para poder tener en promedio de 5 publicaciones por día de los cuatro medios de comunicación que actualmente contamos.

Para que se entienda mejor se mostrará el ciclo del proceso de publicación de noticias manualmente. (Ver imagen 14).

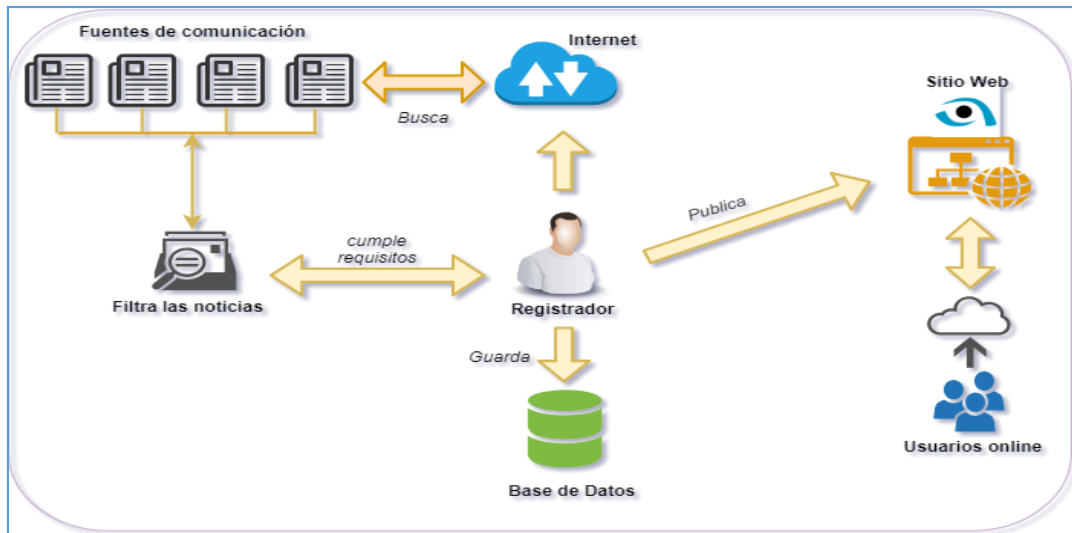


Figura 14: Ciclo del proceso de publicación de noticias

Fuente: Ojotecnologico

Elaboración: propia

La necesidad de estar informados y actualizados sobre noticias tecnológicas en la mayoría de casos para los usuarios se les dificulta o es un problema porque tiene que estar buscando y abriendo una y otra página para encontrar noticias de su interés, en ojotecnologico es un solo sitio que puedan encontrar noticias tecnológicas, y de esta manera pueden ahorrar tiempo.

Se mostrar de qué manera se registraba una noticia manualmente. Para lo cual debemos ir a link : <http://www.ojotecnologico.com/admin/> (Ver imagen 15), donde debemos iniciar sesión con una cuenta de administrador.

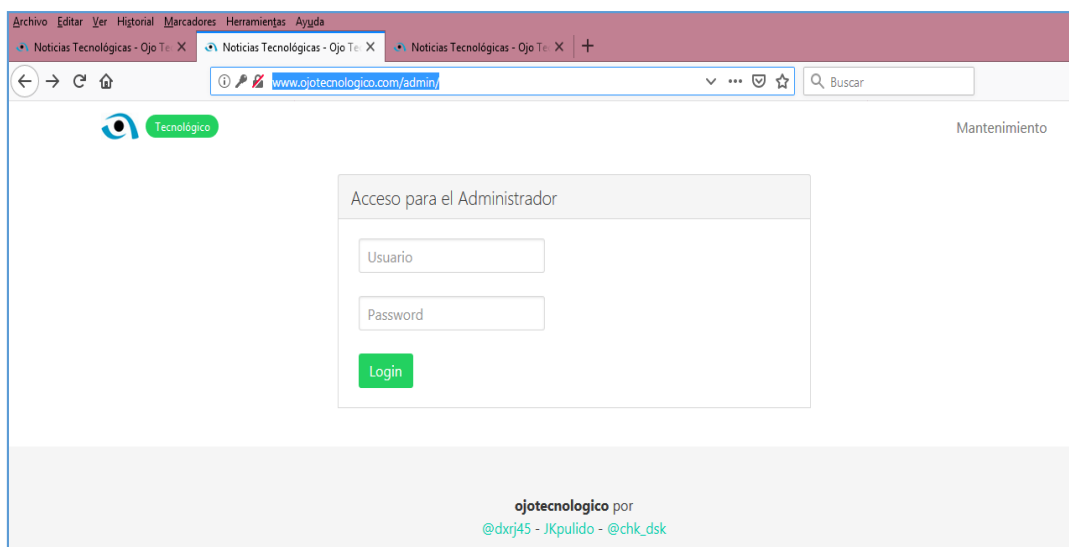


Figura 15: Acceso para el administrador.

Fuente: Ojotecnologico

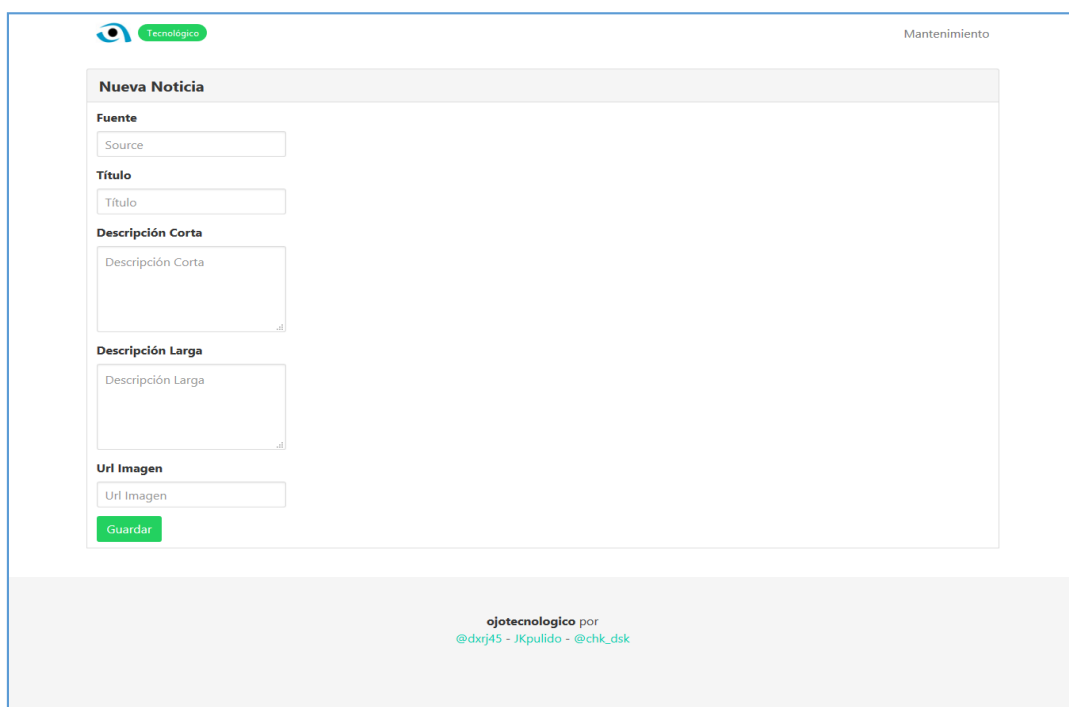
Elaboración: propia

Luego nos mostrará el panel donde llenaremos los campos con la información previamente recolectada de la noticia a guardar.

Los campos requeridos y obligatorios son:

- **Fuente.**-Va el nombre del medio de comunicación.
- **Título.**- Se pone el título de la noticia a guardar.
- **Descripción corta.**- Es un extracto de unos 140 caracteres del texto de la noticia para que se visualizado en la web.
- **Descripción larga.**- Aquí se debe copiar el texto completo de la noticia.
- **URL.**- Aquí se deberá pegar la url de la imagen de la noticia a publicar

Una vez llenados dichos campos se presiona el botón GUARDAR, con esto se da como finalizado la publicación de la noticia. Ver imagen 16 de panel del registro de la noticia.



The image shows a web interface for creating a new news item. At the top left is the 'Ojotecnologico' logo, and at the top right is the text 'Mantenimiento'. The main content area is titled 'Nueva Noticia' and contains several input fields: 'Fuente' (Source), 'Título' (Title), 'Descripción Corta' (Short Description), 'Descripción Larga' (Long Description), and 'Url Imagen' (Image URL). Each field has a placeholder text. Below the 'Url Imagen' field is a green 'Guardar' (Save) button. At the bottom of the page, there is a footer with the text 'ojotecnologico por @dxrj45 - JKpulido - @chk\_dsk'.

Figura 16: Panel de ingreso de nuevas noticias..

Fuente: Ojotecnologico

Elaboración: propia

El resultado es las noticias publicadas en el portal ojotecnologico, como se muestra en la figura 17.

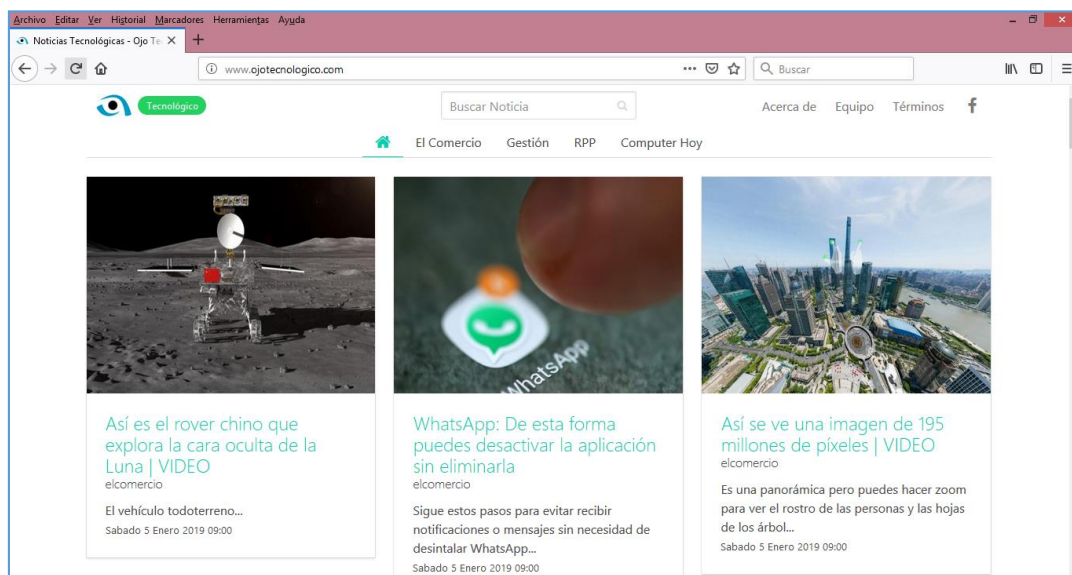


Figura 17: Página principal de ojotecnologico

Fuente: Ojotecnologico

Elaboración: propia

Luego de haber realizado el proceso de publicación de noticias y de acuerdo de nuestra muestra de cantidad de noticias publicadas de la situación actual, a continuación, se detalla cuantas noticias por medio de comunicación fueron publicadas en la segunda semana (7 días) de mayo del 2021, el cual estos datos servirán para el estudio de nuestra investigación. (Ver tabla 12).

Tabla 12  
Situación actual de número de noticias por día

Días / Fuentes	Rpp	Gestión	El Comercio	Computer Hoy	Total
Lunes	2	1	3	1	7
Martes	1	2	3	0	6
Miércoles	2	1	2	3	8
Jueves	1	1	4	3	9
Viernes	1	0	3	4	8
Sábado	2	2	3	3	10
Domingo	1	2	4	2	9

Fuente: Ojotecnologico

Elaboración propia

### 3.1.2 Descripción de la implementación web scraping

La necesidad de estar informados y actualizados sobre noticias tecnológicas en la mayoría de casos para los usuarios se les dificulta o es un problema porque tiene que estar buscando y abriendo una y otra página para encontrar noticias de su interés, por tal motivo se propone la mejora de publicaciones de noticias tecnológicas en el portal web ojotecnologico donde un solo sitio pueda encontrar noticias tecnológicas, y de esta manera pueden ahorrar tiempo.

La aplicación web scraping fue desarrolladas por un integrante de ojotecnologico, para ello se realizó un análisis de lo que se requería para aumentar la cantidad de noticias publicadas diariamente en el portal web.

Nuestro proceso de scraping es capaz de procesar grandes volúmenes de información en un tiempo que podríamos decir relativamente corto el cual está provisto con solidas medidas de seguridad cuando se trabaja con información sensible como sería la protección de datos personales, otra mediad que optamos es que continuamente se pueden hacer cambios estructurales para los procesos de publicación de noticias el cual es prescindible para la mejora continua.

La tecnología empleada para el desarrollo de la aplicación web scraping:

- JavaScript
- Nodejs con express.js
- Rss Reader
- Mongo DB

Se realizó un diagrama de flujo que simplifica y se automatiza el proceso de publicación de noticias. (Ver imagen 18)

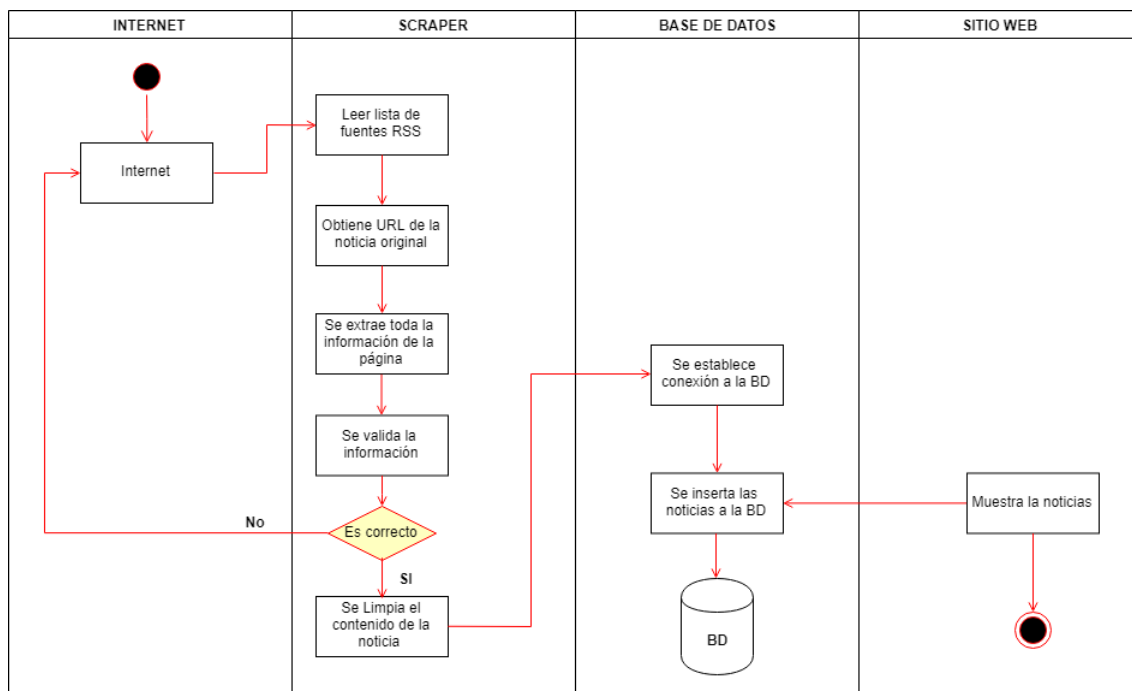


Figura 18: Flujo del proceso de publicación de noticias con web scraping

Fuente: Ojotecnologico

Elaboración: propia

A continuación, indico los pasos del scraper hasta que se guarde la noticia en la base de datos.

- ✓ El scraper lo primero que se hace es leer la fuente de comunicación (El comercio, Gestion, RPP, TICBeat, ComputerHoy, Perú 21 y New YorkTimes).
- ✓ Luego se utiliza el RSS Reader para pasarle la URL de la fuente de comunicación que deseamos obtener la noticia.
- ✓ Se creó un controlador que se encarga de obtener la noticia por medio de la URL de su RSS
- ✓ Como respuesta obtenemos la URL de la noticia original.
- ✓ El modulo se encarga de extraer todo el HTML de la URL
- ✓ Como resultado tenemos todo el contenido de la noticia, (esto implica publicidad, link relacionados a la noticia, Facebook, twitter y entre otras cosas que no es relevante o que no tiene nada que ver con la noticia).



- ✓ Aquí es donde el código desarrollado es el que se encarga de limpiar la noticia, se podría decir que es el primer filtro.
- ✓ Luego un segundo código se encarga de hacer la validación de que la noticia cumpla ciertos requisitos (Que cuente con descripción corta, descripción larga, imagen con formato valido,etc) antes de ser almacena en la BD.
- ✓ Si todo salió bien se realiza una conexión al BD para que sea insertado todas las noticias que fueron validadas exitosamente.
- ✓ Luego el Frond End del portal ojotecnologico se encarga de mostrar la noticia

A continuación, mostramos la arquitectura de todo el proyecto ojotecnologico pero se resaltara la parte del backend Layer, que esta investigación a tratado, que es la mejora del proceso de publicación de noticias con la técnica scraping. (Ver imagen 19)

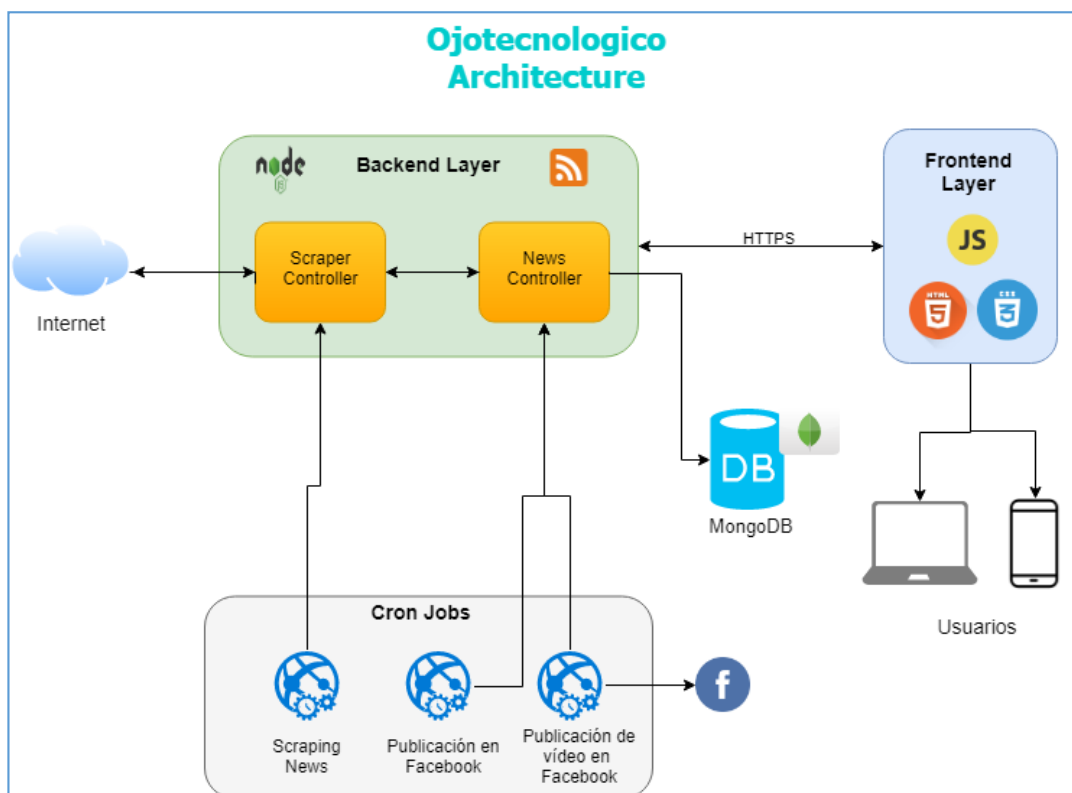


Figura 19: Arquitectura del portal ojotecnologico.

Fuente: Ojotecnologico

Elaboración: propia

Como podemos observar en la arquitectura del portal ojotecnologico notamos claramente que el Backend layer es el entorno de ejecución basado en Node.js donde el

código JavaScript se ejecuta utilizando un patrón de MVC. Esto nos permite tener controladores para acciones específicas.

- **Scraper Controller.**- Se encarga de scrapear la noticia tecnológica, para luego limpiarla y validar que cumpla con los requisitos establecidos para que sea guardada en la BD.
- **New Controller.**- Se encarga de consultar y de guardar las noticias en la BD, actualiza las que fueron publicadas.
- **El Frond Layer.**- Es la página web y este se conecta al backend (NewController) para consultar las noticias y de esta manera ser mostradas en el home de la página web.

Como resultado tenemos las noticias publicadas en el portal ojotecnologico. Como podemos ver en la imagen 20.

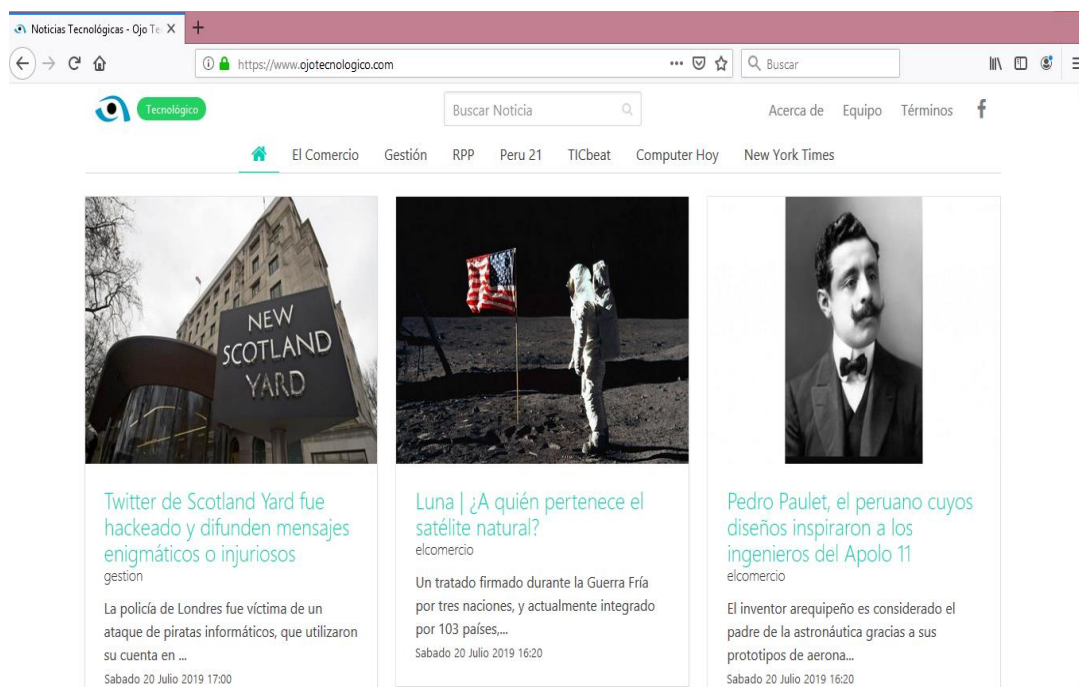


Figura 20:Portal web ojotecnologico.

Fuente: Ojotecnologico

Elaboración: propia

Luego de aplicar la técnica scraping para la mejora del proceso de publicación de noticias en el portal ojotecnologico, mostraremos la segunda semana (7 días) de junio del 2019 la cantidad de noticias por día de cada medio de comunicación. Estos datos se usarán para estudio de nuestra investigación. (Ver tabla 13).

Tabla 13  
Cantidad de noticias diarias luego de aplicar web Scraping

Días / Fuentes	Rpp	Gestión	El Comercio	Computer Hoy	Perú 21	TicBeat	New Yor Times	Total
Lunes	3	1	3	2	2	8	8	27
Martes	3	8	2	6	6	8	4	37
Miércoles	9	3	1	9	6	4	8	40
Jueves	8	6	5	7	6	2	6	40
Viernes	3	3	3	5	6	7	1	28
Sábado	1	1	5	2	4	7	9	29
Domingo	10	2	8	7	3	3	9	42

Fuente: Ojotecnologico  
Elaboración propia

El análisis de la situación actual de la satisfacción de los lectores se ve afectada por las diferentes deficiencias que hay en el proceso de publicar una noticia en el portal web ojotecnologico.

Los lectores no están encontrando noticias actuales y se encuentra con poca variedad de noticias con los medios de comunicación que actualmente tiene ojotecnologico, otro problema es que cuando leen una noticia no está la información completa, aunque este último punto los lectores pueden optar con dar clic en la fuente de origen de la noticia el cual es conducirá a leer la noticia completa.

Para medir el nivel de satisfacción de los lectores del portal ojotecnologico, se preparó una encuesta que tiene como objetivo tener información de la opinión de los lectores.

- El universo de lectores es la fan Page de Facebook del portal web ojotecnologico que actualmente cuenta con 800 seguidores.
- El tamaño de la muestra para la encuesta se seleccionó 15 personas de manera aleatoria.
- Técnica de recolección de datos, cuestionario online.

- Fecha que se realizó la encuesta, 12 de mayo del 2021.

Luego de implementar la web scraping al portal ojotecnologico para mejorar el proceso de publicación de noticias, con el cual se puede contar con más fuentes de comunicación y aumentar el número de noticias diarias y actualizadas. Esta implementación se describió en la hipótesis 1.

A continuación, mostramos como está la web actualmente con seis (6) medios de comunicación cuatro (4) de ellos son medios de comunicación nacionales y tres (3) son medios internacionales, cabe señal que un medio las noticias este inglés. Con esto intentamos tener contenido variado para diferentes tipos de lectores. (Ver imagen 21).

The image shows a screenshot of a news portal website. At the top, there is a search bar labeled "Buscar Noticia" and navigation links for "Acerca de", "Equipo", and "Términos". Below the search bar, there are several category links: "El Comercio", "Gestión", "RPP", "Peru 21", "TICbeat", "Computer Hoy", and "New York Times".

The main content area is a grid of article cards. Each card features a thumbnail image, a headline, a sub-headline, a brief description, and a timestamp. The articles include:

- Twitter de Scotland Yard fue hackeado y difunden mensajes enigmáticos o injuriosos** (elcomercio) - La policía de Londres fue víctima de un ataque de piratas informáticos, que utilizaron su cuenta en... Sabado 20 Julio 2019 17:00
- Luna | ¿A quién pertenece el satélite natural?** (elcomercio) - Un tratado firmado durante la Guerra Fría por tres naciones, y actualmente integrado por 103 países... Sabado 20 Julio 2019 16:20
- Pedro Paulet, el peruano cuyos diseños inspiraron a los ingenieros del Apolo 11** (elcomercio) - El inventor arequipeño es considerado el padre de la astronáutica gracias a sus prototipos de aeronas... Sabado 20 Julio 2019 16:20
- Cómo la imaginación puede ayudarte a cumplir tus metas** (elcomercio) - Según especialistas, la técnica de imaginar tiene resultados positivos innegables y sirve para una m... Sabado 20 Julio 2019 16:20
- Nave Soyuz despegó rumbo a Estación Espacial en homenaje al Apolo 11 | FOTOS** (elcomercio) - La nave rusa Soyuz despegó con ayuda del cohete portador Soyuz-FG y se acoplará a la plataforma orb... Sabado 20 Julio 2019 16:20
- App Ready + Seguridad = Negocio digital que puedes confiar.** (CenturyLink) - El 49% de los líderes digitales priorizan la seguridad de los datos. Las soluciones de conectividad, nube y seguridad de CenturyLink lo ayudan a transformarse ahora. **CONOZCA MÁS**
- Así será el aspecto de Nightwolf en Mortal Kombat 11** (rpp) - Ed Boon, director del título, mostró al luchador, en las redes sociales al c... Sabado 20 Julio 2019 15:29
- Si tienes estas 8 extensiones de Chrome o Firefox te están robando datos, ¡desinstálalas!** (computerhoy) - Muchas extensiones de Chrome o Firefox son directamente malware. Se camuflan como extensiones inocen... Sabado 20 Julio 2019 14:35
- Adiós a Dragonfly: Google cancela el polémico buscador censurado en China** (ticbeat) - Dragonfly ha sido una de las polémicas de Google que más ha sonado durante los últimos meses. No han... Sabado 20 Julio 2019 13:31

At the bottom of the page, there is a footer with a logo and social media handles: "ojotecnologico por @dxj45 - @kpulido - @chk\_dsk".

Figura 21: Portal web luego de implementar el scraping.

Fuente: Ojotecnologico

Elaboración: propia

El haber mejorado el proceso de publicación ha permitido tener siete (7) fuentes de comunicación entre ellas un medio en inglés, cabe indicar que estos medios son de buena reputación y que ya cuenta con un público objetivo.

Para poder implementar la web scraping se tendría que mejorar las características del hosting donde se encuentra alojado el portal web ojotecnologico, y esto se debe porque el procesamiento de información amerita más memoria RAM y va haber mayor tráfico, así mismo debe soportar software como MongoDB, Nodejs y Rss Reader con el cual trabaja la web scraping.

A continuación, mostraremos un cuadro comparativo de las características que contaba el hosting de ojotecnologico y luego se tuvo que mejorar para poder implementar la web scraping. Y los datos de costos nos servirá para utilizarlos en nuestra investigación. (Ver tabla 14).

Tabla 14  
Cantidad de noticias diarias luego de aplicar web Scraping

Características	Pre	Post
Espacio en Disco	10 gb	20 GB
Memoria RAM	500 Mb	1GB
Transferencia Mensual	500GB	1000 GB
Copias de seguridad	no	no
Certificado SSL	no	Si
Acceso a logs	Si	Si
Cuentas de correo	1	2
Administrador de ficheros	Si	Si
Accesos FTP	Si	Si
Base de datos MySQL 5.x	Si	Si
Soporte PHP 5.x	Si	Si
Bases de datos PostgreSQL	Si	Si
Soporte MongoDB	No	Si
Nodo.JS 6, 8, 9	No	Si
Javascript	Si	Si
Acceso SSH (Encarcelado)	Si	Si
Cron Jobs	No	Si
Gestión de permisos	No	Si
<b>Total a pagar Mensual</b>	<b>16.45</b>	<b>23.03</b>

Elaboración propia

Fuente: DigitalOcen Inc.

En cuanto a la cantidad de vistas al portal web de noticias tecnológicas ojotecnologico, luego de implementarse la web scraping se ve que ha incrementado considerablemente, el cual se puede ver en el informe obtenido de Google Analytics, en donde se compara el mes de estudio de la investigación que es abril del 2021 con abril del 2022, obteniendo el total de audiencia que paso de 369 vistas a un total de 1426 visitas, el cual se estima que cada vez aumente y esto se debe que encontrarán variedad de noticias tecnológicas los lectores.

Podemos notar que aumentado nuevos usuarios y nuevas sesiones el cual se positivo para el portal web ojotecnologico. (Ver imagen 22).

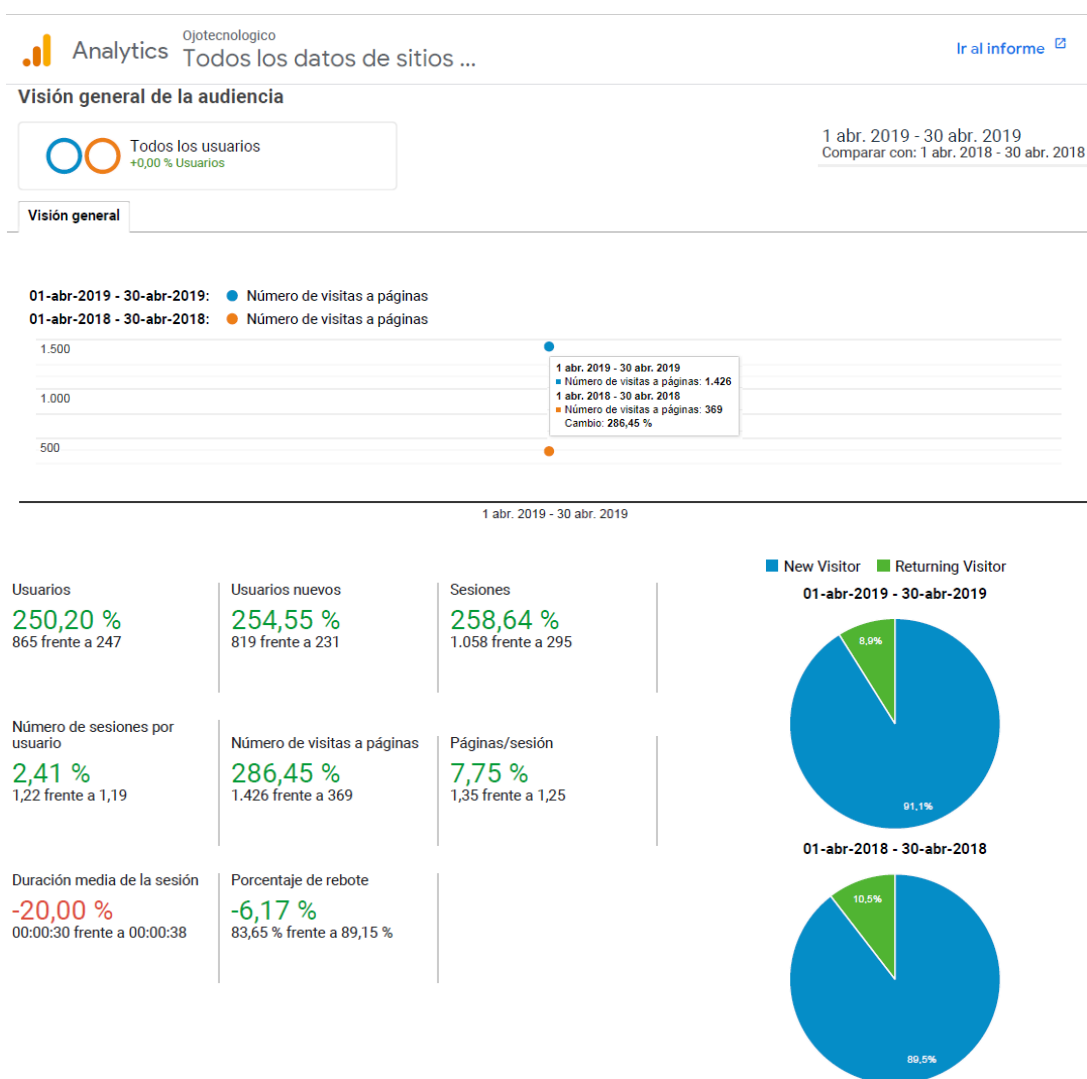


Figura 22: Comparación de usuarios y sesiones nuevas

Fuente: Google Analytics

Elaboración: propia

Para saber el nivel de usabilidad del portal web ojotecnologico se consultó páginas especializadas que sirven como herramientas para saber velocidad de carga, análisis de contenido, si es adaptativo a cualquier dispositivo, los resultados son favorables como puede observarse a continuación (Ver imagen 23).

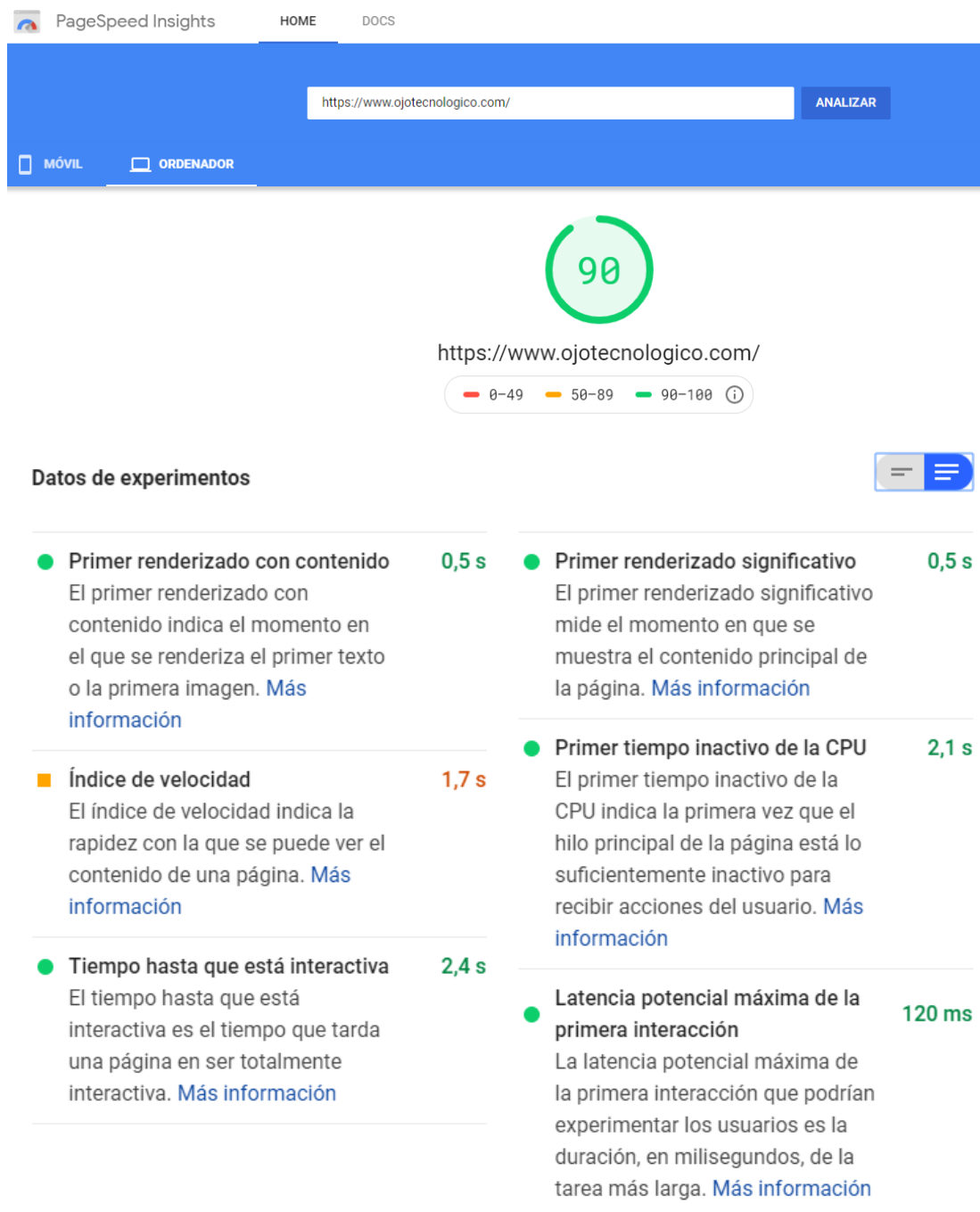


Figura 23: Puntaje de Aplicación web progresiva, Rendimiento, Accesibilidad

Fuente: <https://developers.google.com>

Elaboración: propia



GTmetrix pagina especializada en en medir el rendimiento de de una web bazados en indicadores de velocidad. Y para el portal web ojotecnologico lo resultados son lo siguientes: (ver Imagen 24)

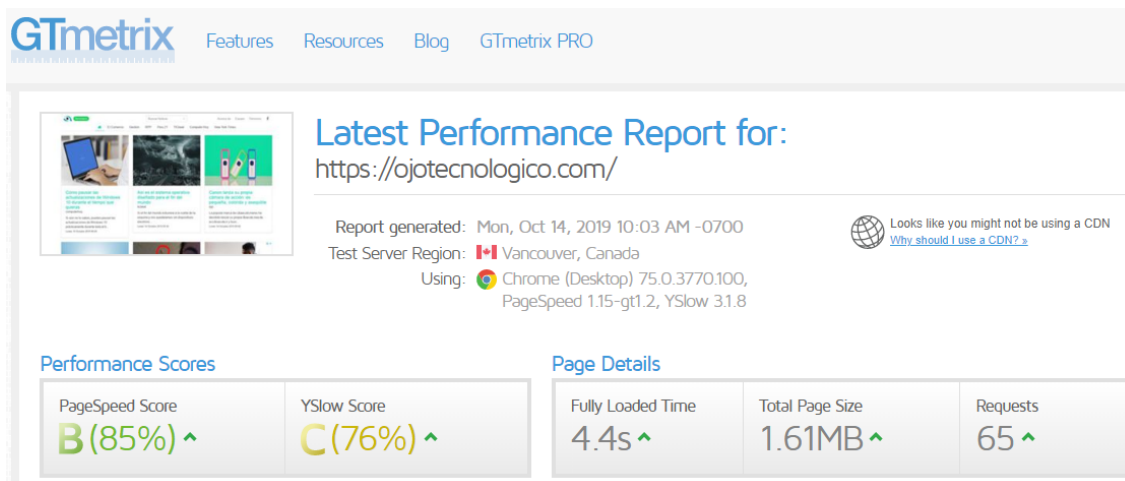


Figura 24: Informe de rendimiento de ojotecnologico.

Fuente: <https://gtmetrix.com>

Elaboración: propia

Aquí la evidencia de como el portal web ojotecnologico es adaptativo a cual quier dispositivo para que los lectores puedan acceder a las noticias sin ningun inconveniente. Ver imagen 25

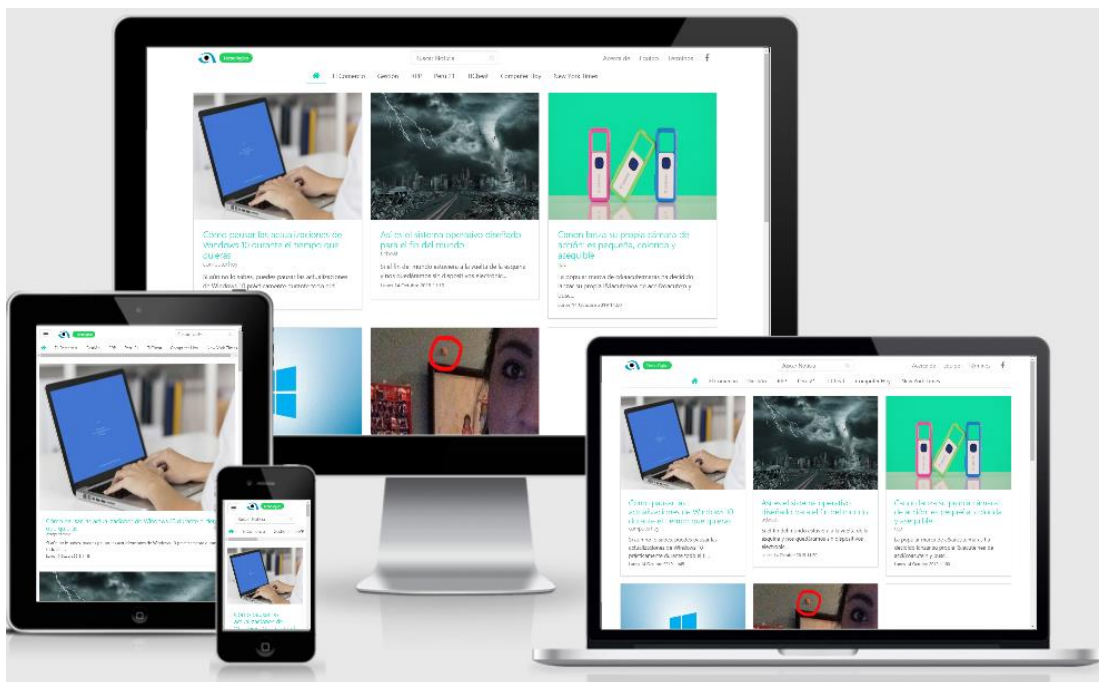


Figura 25: Diseño Responsivo de ojotecnologico.

Fuente: <http://ami.responsivedesign.is>

Elaboración: propia

### 3.1.3 Análisis e interpretación de resultados de web scraping

Iniciamos valorando la variable de web scraping se utilizó el instrumento de recolección de datos elaborando un formulario con preguntas validado y confiable, que está conformado por 12 ítems o preguntas que estuvieron destinados a recoger la percepción sobre las dimensiones noticias publicadas por día (04 ítems), Visitas por día (04 ítems) y usabilidad (04 ítems).

Los ítems de las preguntas se elaboraron en relación a la matriz de operacionalización de variable, teniendo como premisa sus dimensiones, cada ítem es de tipo cerrado y utiliza la escala de Likert para valorar la apreciación sobre cada uno de ellos en una escala de 1 a 5 siguiendo la siguiente escala de valoración:

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni desacuerdo
- De acuerdo
- Totalmente de acuerdo

El total de encuestados permitió crear una base de datos de respuestas sobre el que se ejecutó el análisis estadístico correspondiente con el software SPSS, a continuación, se formularon ordenadamente de acuerdo a las dimensiones e indicadores de la variable web scraping

#### 3.1.3.1 Sobre la dimensión Noticias publicadas por día

Tabla 15  
Resultado Descriptivo sobre la pregunta uno 1 (Pre Test)

#### P1. ¿Según su criterio la cantidad de noticias publicadas por día es la correcta?

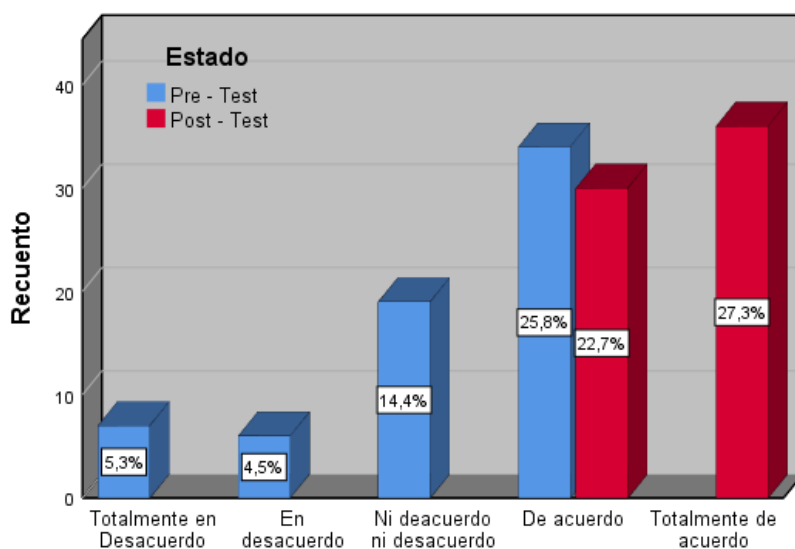
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	7	10,6	10,6	10,6
	En desacuerdo	6	9,1	9,1	19,7
	Ni de acuerdo ni desacuerdo	19	28,8	28,8	48,5
	De acuerdo	34	51,5	51,5	100,0
	Total	66	100,0	100,0	

Tabla 16  
Resultado Descriptivo sobre la pregunta uno 1 (Post Test)

**P1. ¿Según su criterio la cantidad de noticias publicadas por día es la correcta?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	30	45,5	45,5	45,5
	Totalmente de acuerdo	36	54,5	54,5	100,0
	Total	66	100,0	100,0	

**P1. ¿Según su criterio la cantidad de noticias publicadas por día es la correcta? por Estado**



**P1.**

Figura 26: Resultados descriptivos sobre la pregunta 1 (Pre y Post)  
Fuente: Base de datos de la encuesta

**Interpretación:** Podemos observar que la cantidad de noticias por día es importante para el lector, luego de que se haya implementado la web scraping, esto facilita a los lectores que encuentren variedad de noticias todos los días. Adicional una mejora considerable que la frecuencia de 28 % de usuarios que estaba ni de acuerdo y ni desacuerdo en la prueba post bajo a 0%, esto nos indica que los lectores están conformes con la cantidad de noticias diarias.

Tabla 17  
Resultado Descriptivo sobre la pregunta 2 (Pre Test)

**P2. ¿A su criterio las fuentes de comunicación que se muestran actualmente es suficiente o se debería aumentar más?**

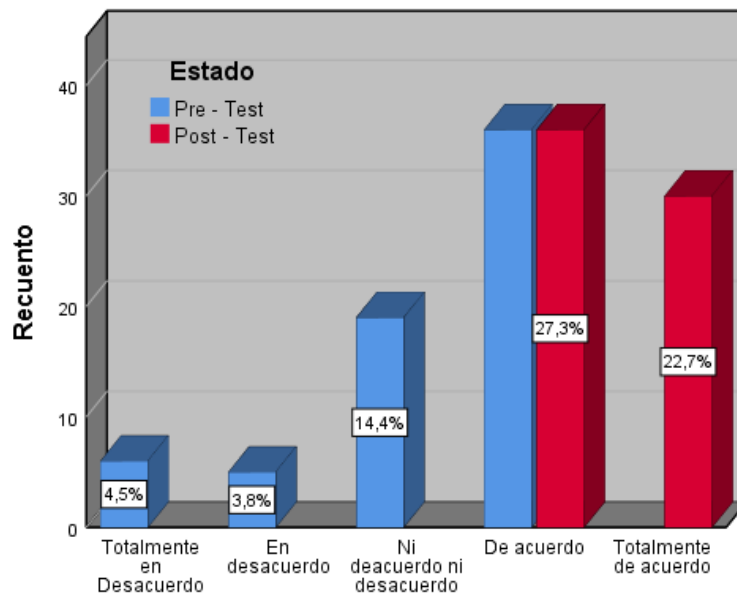
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	6	9,1	9,1	9,1
	En desacuerdo	5	7,6	7,6	16,7
	Ni de acuerdo ni desacuerdo	19	28,8	28,8	45,5
	De acuerdo	36	54,5	54,5	100,0
	Total	66	100,0	100,0	

Tabla 18  
Resultado Descriptivo sobre la pregunta 2 (Post Test)

**P2. ¿A su criterio las fuentes de comunicación que se muestran actualmente es suficiente o se debería aumentar más?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	36	54,5	54,5	54,5
	Totalmente de acuerdo	30	45,5	45,5	100,0
	Total	66	100,0	100,0	

**P2. ¿A su criterio las fuentes de comunicación que se muestran actualmente es suficiente o se debería aumentar más? por Estado**



**P2.**

Figura 27: Resultados descriptivos sobre la pregunta 2 (Pre y Post)  
Fuente: Base de datos de la encuesta

**Interpretación:** Con respecto a la pregunta n° 2 podemos observar que los lectores están interesados que haya varias fuentes de comunicación donde puedan consultar noticias de su interés, y luego de aplicar la web scraping tenemos un 22.7% que si está totalmente de acuerdo que se incremente las fuentes de comunicación.

Tabla 19  
Resultado Descriptivo sobre la pregunta 3 (Pre Test)

**P3. ¿Es importante saber que fuente de comunicación pública más noticias por día?**

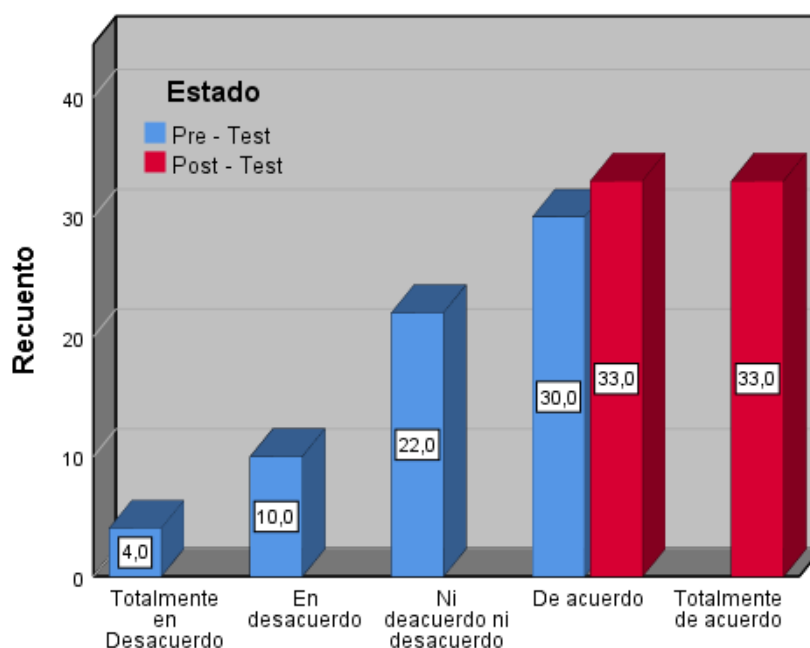
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	4	6,1	6,1	6,1
	En desacuerdo	10	15,2	15,2	21,2
	Ni de acuerdo ni desacuerdo	22	33,3	33,3	54,5
	De acuerdo	30	45,5	45,5	100,0
	Total	66	100,0	100,0	

Tabla 20  
Resultado Descriptivo sobre la pregunta 3 (Post Test)

**P3. ¿Es importante saber que fuente de comunicación pública más noticias por día?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	33	50,0	50,0	50,0
	Totalmente de acuerdo	33	50,0	50,0	100,0
	Total	66	100,0	100,0	

**P3. ¿Es importante saber que fuente de comunicación publica más noticias por día? por Estado**



**P3.**

Figura 28: Resultados descriptivos sobre la pregunta 3 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Notamos claramente que luego de la implementación de la web scraping aumento un 66 % de que se contabilice que fuente de información es el que brinda más noticias, y con esto también se puede analizar que fuente es que la menos le interesa a los lectores y se podría tomar la decisión de reemplazarlo por otra fuente.

Tabla 21

Resultado Descriptivos sobre la pregunta 4 (Pre Test)

**P4. ¿Está de acuerdo que solo encuentra noticias únicas y no duplicadas?**

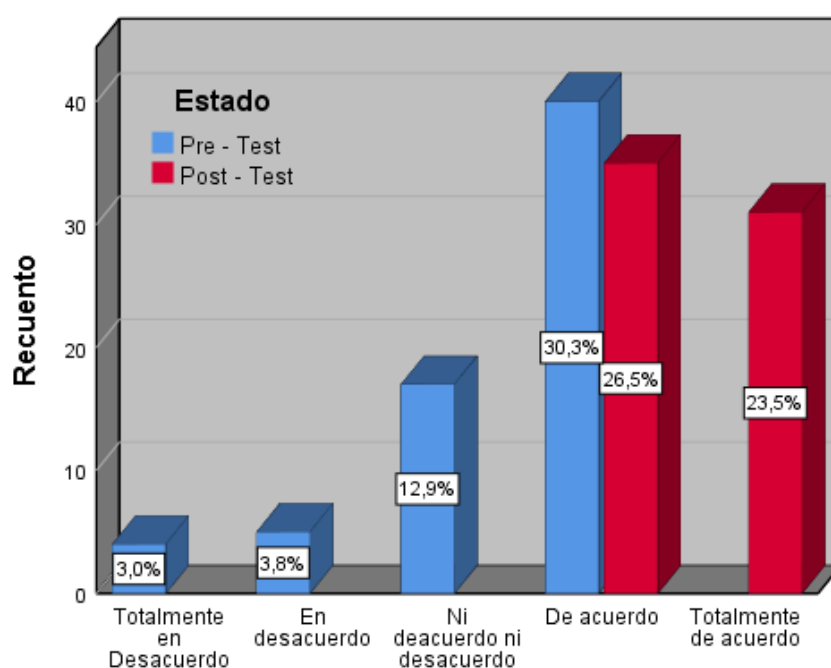
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	4	6,1	6,1	6,1
	En desacuerdo	5	7,6	7,6	13,6
	Ni de acuerdo ni desacuerdo	17	25,8	25,8	39,4
	De acuerdo	40	60,6	60,6	100,0
	Total	66	100,0	100,0	

Tabla 22  
Resultados Descriptivo sobre la pregunta 4 (Post Test)

**P4. ¿Está de acuerdo que solo encuentra noticias únicas y no duplicadas?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	35	53,0	53,0	53,0
	Totalmente de acuerdo	31	47,0	47,0	100,0
	Total	66	100,0	100,0	

**P4. ¿Está de acuerdo que solo encuentra noticias únicas y no duplicadas?  
por Estado**



**P4.**

Figura 29: Resultados descriptivos sobre la pregunta 4 (Pre y Post)  
Fuente: Base de datos de la encuesta

**Interpretación:** El análisis que podemos hacer de la pregunta cuatro es que los lectores quieren siempre noticias únicas y no duplicadas, el cual la web scraping garantiza ya que las noticias publicadas tienen que cumplir ciertos filtros.

### 3.1.3.2 Sobre la dimensión de Visitas por día

Tabla 23  
Resultado Descriptivo sobre la pregunta 5 (Pre Test)

#### P5. ¿A su opinión se ve un incremento de visitas al portal web?

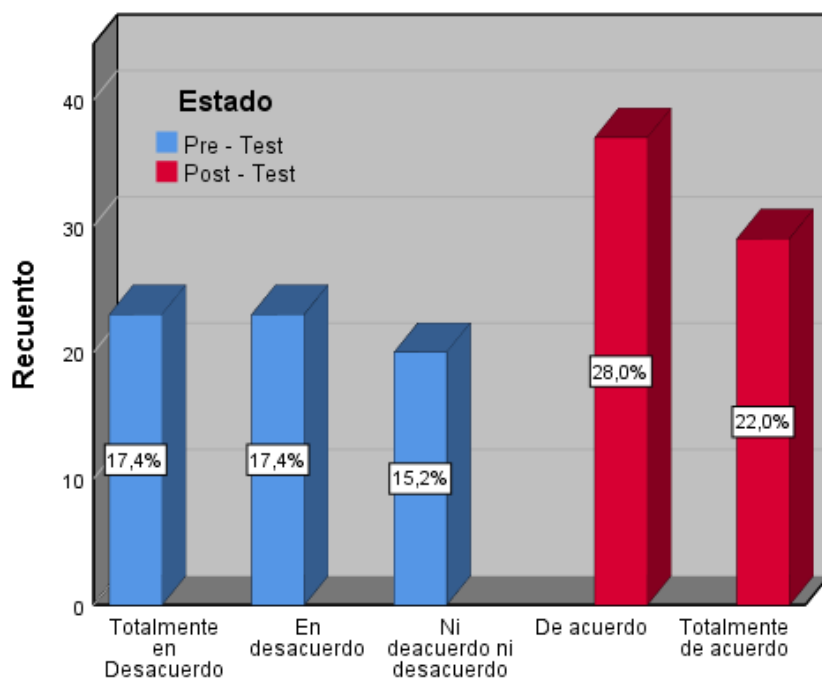
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	23	34,8	34,8	34,8
	En desacuerdo	23	34,8	34,8	69,7
	Ni de acuerdo ni desacuerdo	20	30,3	30,3	100,0
	Total	66	100,0	100,0	

Tabla 24  
Resultado Descriptivo sobre la pregunta 5 (Post Test)

#### P5. ¿A su opinión se ve un incremento de visitas al portal web?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	37	56,1	56,1	56,1
	Totalmente de acuerdo	29	43,9	43,9	100,0
	Total	66	100,0	100,0	

#### P5. ¿A su opinión se ve un incremento de visitas al portal web? por Estado



P5.

Figura 30: Resultados descriptivos sobre la pregunta 5 (Pre y Post)

Fuente: Base de datos de la encuesta



**Interpretación:** Luego de implementar la web scraping la cantidad de vistas al portal ojetecnologico ha tenido una mejora considerable, y esto se debe para la cantidad de noticias actualizadas y fuentes de comunicación importantes tanto nacionales como internacionales.

Tabla 25  
Resultados Descriptivos sobre la pregunta 6 (Pre Test)

**P6. ¿A su criterio ha aumentado el número de usuarios recurrentes?**

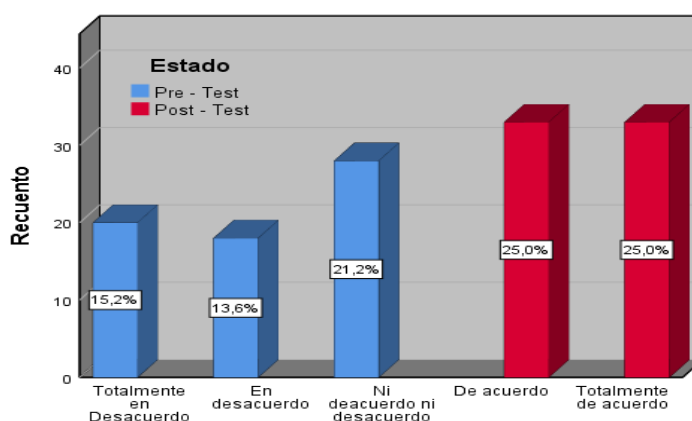
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	20	30,3	30,3	30,3
	En desacuerdo	18	27,3	27,3	57,6
	Ni de acuerdo ni desacuerdo	28	42,4	42,4	100,0
	Total	66	100,0	100,0	

Tabla 26  
Resultado Descriptivo sobre la pregunta 6 (Post Test)

**P6. ¿A su criterio ha aumentado el número de usuarios recurrentes?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	33	50,0	50,0	50,0
	Totalmente de acuerdo	33	50,0	50,0	100,0
	Total	66	100,0	100,0	

**P6. ¿A su criterio ha aumentado el número de usuarios recurrentes? por Estado**



P6.

Figura 31: Resultados descriptivos sobre la pregunta 6 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Observamos que los usuarios recurrentes han aumentado considerablemente luego de que se implementara la web scraping, esto es significativo

debido que la pagina va ganando más lectores cada día, y esto conlleva a realizar mejores constantes en el portal web ojotecnologico.

Tabla 27  
Resultado Descriptivo sobre la pregunta 7 (Pre Test)

**P7. ¿La automatización de la publicación de noticias ha tenido un impacto positivo en la afluencia de usuarios?**

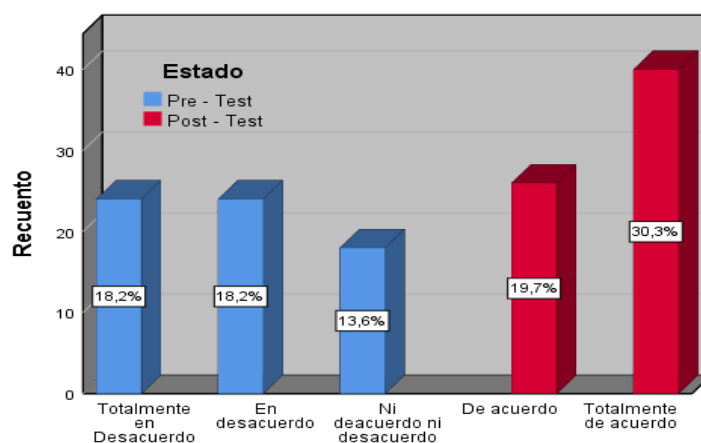
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	24	36,4	36,4	36,4
	En desacuerdo	24	36,4	36,4	72,7
	Ni de acuerdo ni desacuerdo	18	27,3	27,3	100,0
	Total	66	100,0	100,0	

Tabla 28  
Resultado Descriptivo sobre la pregunta 7 (Post Test)

**P7. ¿La automatización de la publicación de noticias ha tenido un impacto positivo en la afluencia de usuarios?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	26	39,4	39,4	39,4
	Totalmente de acuerdo	40	60,6	60,6	100,0
	Total	66	100,0	100,0	

**P7. ¿La automatización de la publicación de noticias ha tenido un impacto positivo en la afluencia de usuarios? por Estado**



P7.

Figura 32: Resultados descriptivos sobre la pregunta 7 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Podemos observar que la automatización de publicación de noticias impacta positivamente en que nuevos lectores puedan visitar nuestra página, ha

aumentado en más de 50% que las personas en de acuerdo con este tipo de propuesta para la mejora del portal ojotecnologico.

Tabla 29  
Resultado Descriptivo sobre la pregunta 8 (Pre Test)

**P8. ¿Generar informes personalizadas para llevar el control de vistas diarias?**

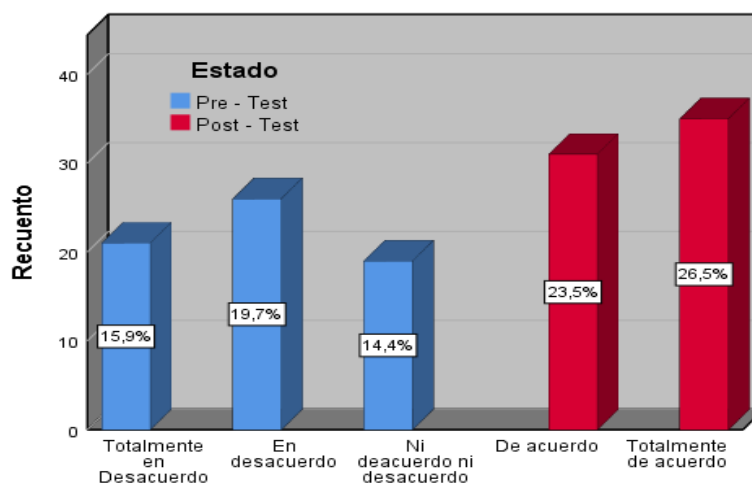
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	21	31,8	31,8	31,8
	En desacuerdo	26	39,4	39,4	71,2
	Ni de acuerdo ni desacuerdo	19	28,8	28,8	100,0
	Total	66	100,0	100,0	

Tabla 30  
Resultado Descriptivo sobre la pregunta 8 (Post Test)

**P8. ¿Generar informes personalizadas para llevar el control de vistas diarias?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	31	47,0	47,0	47,0
	Totalmente de acuerdo	35	53,0	53,0	100,0
	Total	66	100,0	100,0	

**P8. ¿Generar informes personalizadas para llevar el control de vistas diarias?  
por Estado**



P8.

Figura 33: Resultados descriptivos sobre la pregunta 8 (Pre y Post)  
Fuente: Base de datos de la encuesta

**Interpretación:** Podemos notar que es muy importante llevar un control de visitas con el cual se puede saber qué día hubo más visitas al portal ojotecnologico, y hacer un

análisis si hay alguna noticia que tuvo mayor impacto entre los lectores y esto sirve para poder medir que fuente de comunicación es el más leído.

### 3.1.3.3 Sobre la dimensión de Usabilidad

Tabla 31  
Resultado Descriptivo sobre la pregunta 9 (Pre Test)

#### P9. ¿La mayoría de personas aprendería a utilizar nuestra web de noticias tecnológicas?

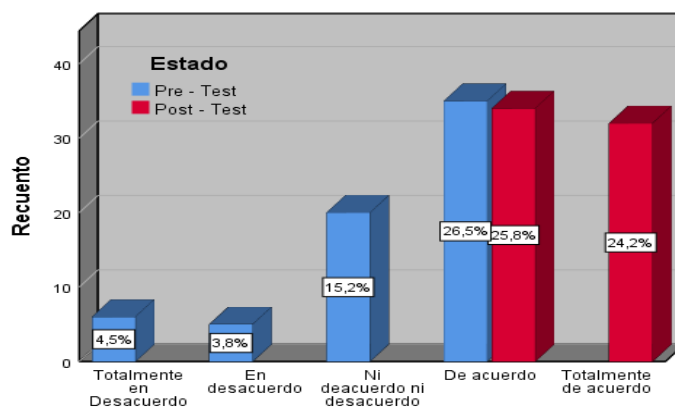
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	6	9,1	9,1	9,1
	En desacuerdo	5	7,6	7,6	16,7
	Ni de acuerdo ni desacuerdo	20	30,3	30,3	47,0
	De acuerdo	35	53,0	53,0	100,0
	Total	66	100,0	100,0	

Tabla 32  
Resultado Descriptivo sobre la pregunta 9 (Post Test)

#### P9. ¿La mayoría de personas aprendería a utilizar nuestra web de noticias tecnológicas?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	34	51,5	51,5	51,5
	Totalmente de acuerdo	32	48,5	48,5	100,0
	Total	66	100,0	100,0	

#### P9. ¿La mayoría de personas aprendería a utilizar nuestra web de noticias tecnológicas? por Estado



P9.

Figura 34: Resultados descriptivos sobre la pregunta 9 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** La usabilidad de nuestro portal web es muy fácil de aprender y la encuesta lo respalda la gran mayoría indica que está de acuerdo, esto permite que las noticias sean fácil de encontrar y de leer, esto seda tanto en web como navegación móvil.

Tabla 33  
Resultado Descriptivo sobre la pregunta 10 (Pre Test)

**P10. ¿Se siente seguro al navegar por nuestra web?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	2	3,0	3,0	3,0
	En desacuerdo	5	7,6	7,6	10,6
	Ni de acuerdo ni desacuerdo	19	28,8	28,8	39,4
	De acuerdo	40	60,6	60,6	100,0
	Total	66	100,0	100,0	

Tabla 34  
Resultado Descriptivo sobre la pregunta 10 (Post Test)

**P10. ¿Se siente seguro al navegar por nuestra web?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	29	43,9	43,9	43,9
	Totalmente de acuerdo	37	56,1	56,1	100,0
	Total	66	100,0	100,0	

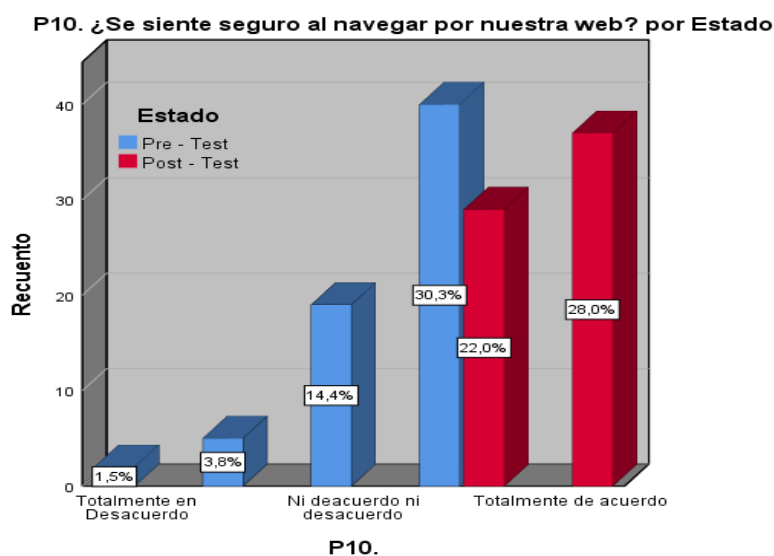


Figura 35: Resultados descriptivos sobre la pregunta 10 (Pre y Post)  
Fuente: Base de datos de la encuesta

**Interpretación:** Podemos observar que antes de implementar la web scraping los lectores se sentían seguros de navegar por la web y esto incremento luego de

implementar la web scraping tal y como podemos apreciar en el gráfico. Y esto un punto a favor del portal web ojotecnologico por que va a tener buenas referencias en cuento a seguridad.

Tabla 35  
Resultado Descriptivo sobre la pregunta 11 (Pre Test)

**P11. ¿Cree usted que es bastante fácil navegar vía web y dispositivo móvil?**

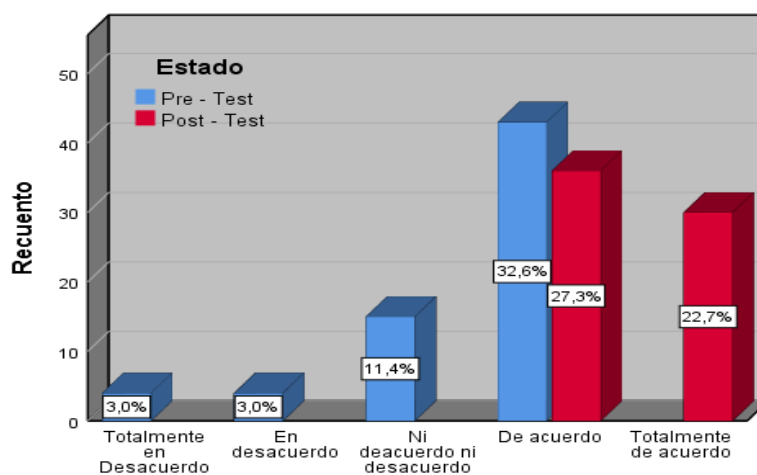
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	4	6,1	6,1	6,1
	En desacuerdo	4	6,1	6,1	12,1
	Ni de acuerdo ni desacuerdo	15	22,7	22,7	34,8
	De acuerdo	43	65,2	65,2	100,0
	Total	66	100,0	100,0	

Tabla 36  
Resultados Descriptivos sobre la pregunta 11 (Post Test)

**P11. ¿Cree usted que es bastante fácil navegar vía web y dispositivo móvil?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	36	54,5	54,5	54,5
	Totalmente de acuerdo	30	45,5	45,5	100,0
	Total	66	100,0	100,0	

**P11. ¿Cree usted que es bastante fácil navegar vía web y dispositivo móvil?  
por Estado**



P11.

Figura 36: Resultados descriptivos sobre la pregunta 11 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Podemos observar que la navegación del portal ojotecnologico es bastante fácil y adaptable a los diferentes dispositivos que hoy en di hay para acceder a la información, y así lo manifiestan nuestros encuestados para nuestra investigación.

Tabla 37

Resultado Descriptivo sobre la pregunta 12 (Pre Test)

**P12. ¿Cree que las diferentes funciones de nuestra web se encuentran muy bien integradas?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	6	9,1	9,1	9,1
	En desacuerdo	1	1,5	1,5	10,6
	Ni de acuerdo ni desacuerdo	24	36,4	36,4	47,0
	De acuerdo	35	53,0	53,0	100,0
	Total	66	100,0	100,0	

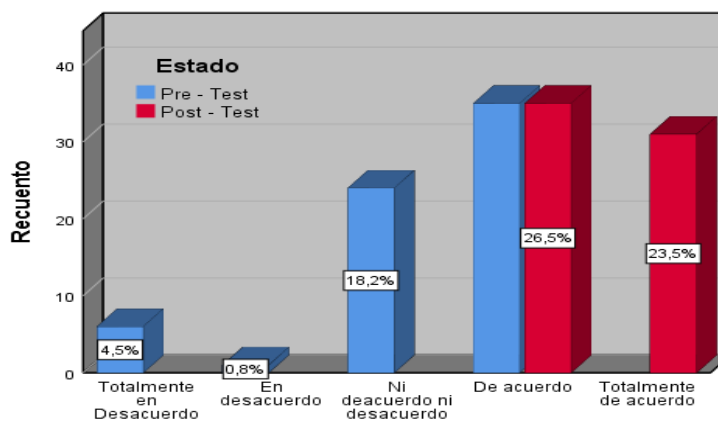
Tabla 38

Resultado Descriptivo sobre la pregunta 12 (Post Test)

**P12. ¿Cree que las diferentes funciones de nuestra web se encuentran muy bien integradas?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	35	53,0	53,0	53,0
	Totalmente de acuerdo	31	47,0	47,0	100,0
	Total	66	100,0	100,0	

**P12. ¿Cree que las diferentes funciones de nuestra web se encuentran muy bien integradas? por Estado**



P12.

Figura 37: Resultados descriptivos sobre la pregunta 12 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Estos resultados nos indican que los usuarios utilizan las funciones del portal ojotecnologico por iniciativa propia el cual conlleva que el lector se sienta cómodo al momento de revisar noticias tecnológicas. Y luego de la implementación de la web scraping aumento la cantidad de fuentes y noticias y el portal web mantiene las funciones que hace más fácil la navegación.

### 3.1.4 Análisis e interpretación de resultados de publicación de noticias

#### 3.1.4.1 Sobre la dimensión de Satisfacción

Tabla 39  
Resultado Descriptivo sobre la pregunta 13 (Pre Test)

**P13. ¿El aspecto de nuestra web es el adecuado para mostrar las noticias tecnológicas?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	24	36,4	36,4	36,4
	En desacuerdo	22	33,3	33,3	69,7
	Ni de acuerdo ni desacuerdo	20	30,3	30,3	100,0
	Total	66	100,0	100,0	

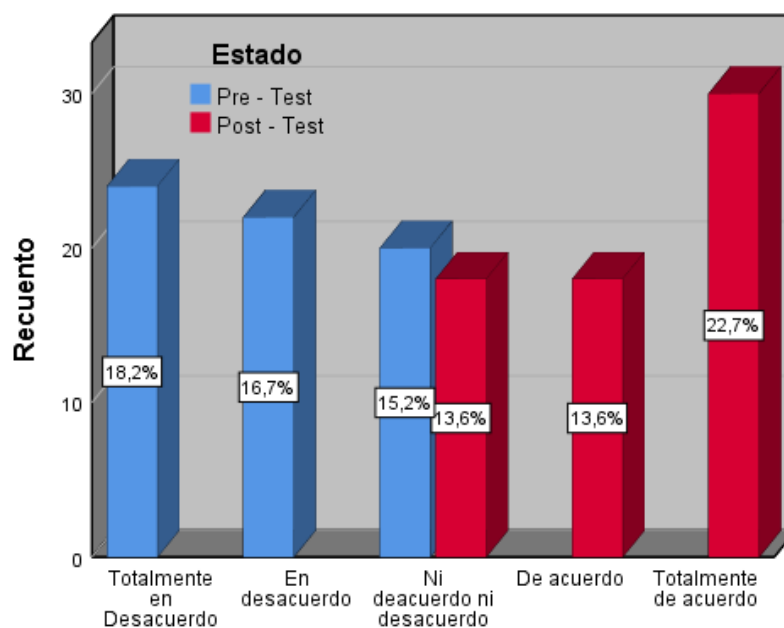
Tabla 40  
Resultado Descriptivo sobre la pregunta 13 (Post Test)

**P13. ¿El aspecto de nuestra web es el adecuado para mostrar las noticias tecnológicas?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Ni de acuerdo ni desacuerdo	18	27,3	27,3	27,3
	De acuerdo	18	27,3	27,3	54,5
	Totalmente de acuerdo	30	45,5	45,5	100,0
	Total	66	100,0	100,0	



**P13. ¿El aspecto de nuestra web es el adecuado para mostrar las noticias tecnológicas? por Estado**



**P13.**

Figura 38: Resultados descriptivos sobre la pregunta 13 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Antes que se implementara el web scraping los lectores no lo veían con un buen aspecto de nuestra web, quizás por las pocas noticias publicadas, ya luego de la automatización de publicación de noticias, es que se realizó mejoras de usabilidad hacia la web el cual vemos que los encuestados indican que el aspecto de favorable.

Tabla 41

Resultado Descriptivo sobre la pregunta 14 (Pre Test)

**P14. ¿La página web transmite la idea del proyecto?**

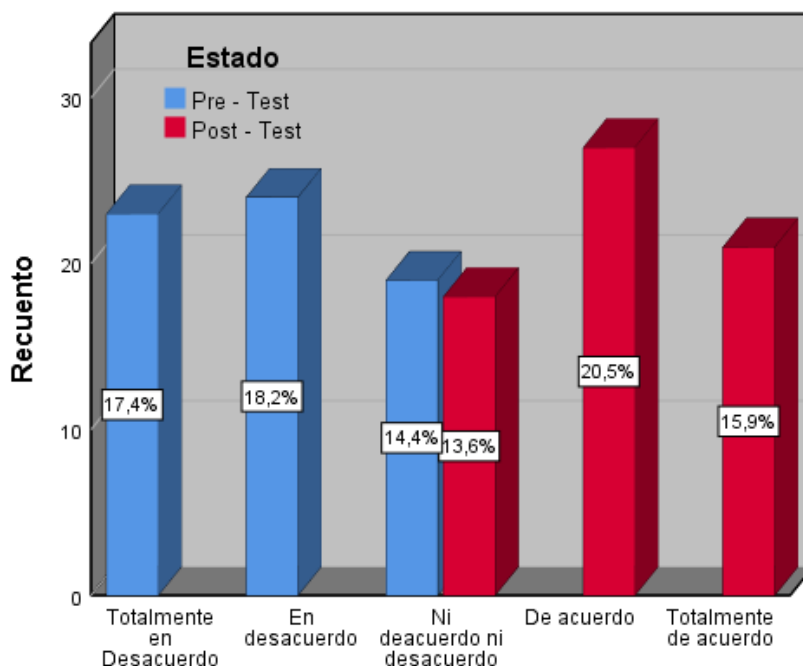
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	23	34,8	34,8	34,8
	En desacuerdo	24	36,4	36,4	71,2
	Ni de acuerdo ni desacuerdo	19	28,8	28,8	100,0
	Total	66	100,0	100,0	

Tabla 42  
Resultado Descriptivo sobre la pregunta 14 (Post Test)

**P14. ¿La página web transmite la idea del proyecto?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Ni de acuerdo ni desacuerdo	18	27,3	27,3	27,3
	De acuerdo	27	40,9	40,9	68,2
	Totalmente de acuerdo	21	31,8	31,8	100,0
	Total	66	100,0	100,0	

**P14. ¿La página web transmite la idea del proyecto? por Estado**



**P14.**

Figura 39: Resultados descriptivos sobre la pregunta 14 (Pre y Post)  
Fuente: Base de datos de la encuesta

**Interpretación:** El portal de noticias ojotecnologico recién transmite la idea de la tesis luego de que se implemente la web scraping y esto se debe a dos factores importantes que se aumentó las fuentes de comunicación tanto nacionales como internacionales y la cantidad de noticias publicadas diariamente.

Tabla 43  
Resultado Descriptivo sobre la pregunta 15 (Pre Test)

**P15. ¿El portal ojotecnologico es una web recomendable?**

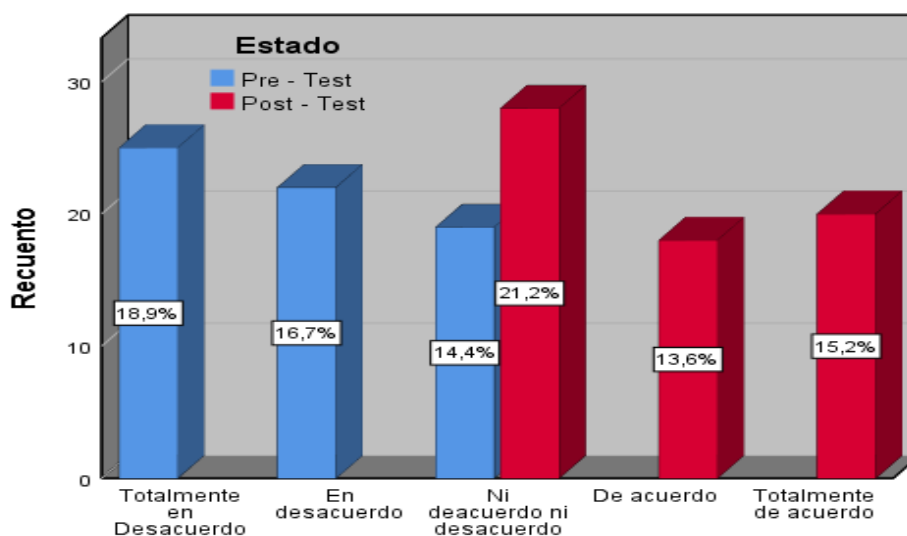
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	25	37,9	37,9	37,9
	En desacuerdo	22	33,3	33,3	71,2
	Ni de acuerdo ni desacuerdo	19	28,8	28,8	100,0
	Total	66	100,0	100,0	

Tabla 44  
Resultado Descriptivo sobre la pregunta 15 (Post Test)

**P15. ¿El portal ojotecnologico es una web recomendable?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Ni de acuerdo ni desacuerdo	28	42,4	42,4	42,4
	De acuerdo	18	27,3	27,3	69,7
	Totalmente de acuerdo	20	30,3	30,3	100,0
	Total	66	100,0	100,0	

**P15. ¿El portal ojotecnologico es una web recomendable? por Estado**



**P15.**

Figura 40: Resultados descriptivos sobre la pregunta 15 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Observamos que el porcentaje de recomendación es mayor al 30% luego de que se implementara la web scraping, esto se debe porque hay noticias

actualizadas y de mucho interés de las personas que son seguidores de las últimas novedades sobre tecnología.

Tabla 45  
Resultado Descriptivo sobre la pregunta 16 (Pre Test)

**P16. ¿Según tu criterio la información que se muestra en Ojotecnología es relevante?**

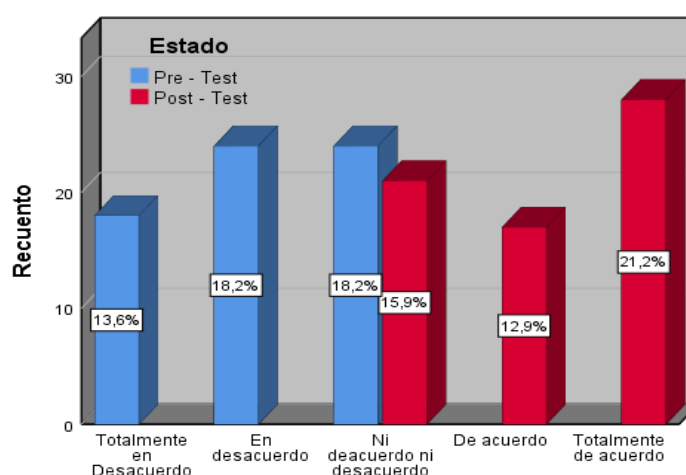
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	18	27,3	27,3	27,3
	En desacuerdo	24	36,4	36,4	63,6
	Ni de acuerdo ni desacuerdo	24	36,4	36,4	100,0
	Total	66	100,0	100,0	

Tabla 46  
Resultado Descriptivo sobre la pregunta 16 (Post Test)

**P16. ¿Según tu criterio la información que se muestra en Ojotecnología es relevante?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Ni de acuerdo ni desacuerdo	21	31,8	31,8	31,8
	De acuerdo	17	25,8	25,8	57,6
	Totalmente de acuerdo	28	42,4	42,4	100,0
	Total	66	100,0	100,0	

**P16. ¿Según tu criterio la información que se muestra en Ojotecnología es relevante? por Estado**



**P16.**

Figura 41: Resultados descriptivos sobre la pregunta 16 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Al analizar los datos podemos decir que al aumentar la cantidad de noticias diarias, el contenido de nuestra web se vuelve relevante por la variedad de

noticias de tecnología, y los lectores tienen una la decisión que noticias leer o excluir al momento de navegar por la web de ojoetecnologico.

### 3.1.4.2 Sobre la dimensión de costos

Tabla 47  
Resultado Descriptivo sobre la pregunta 17 (Pre Test)

#### P17. ¿Resulta más rápido y seguro la automatización de publicación de noticias?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	5	7,6	7,6	7,6
	En desacuerdo	9	13,6	13,6	21,2
	Ni de acuerdo ni desacuerdo	19	28,8	28,8	50,0
	De acuerdo	33	50,0	50,0	100,0
	Total	66	100,0	100,0	

Tabla 48  
Resultados Descriptivos sobre la pregunta 17 (Post Test)

#### P17. ¿Resulta más rápido y seguro la automatización de publicación de noticias?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	33	50,0	50,0	50,0
	Totalmente de acuerdo	33	50,0	50,0	100,0
	Total	66	100,0	100,0	

#### P17. ¿Resulta más rápido y seguro la automatización de publicación de noticias? por Estado

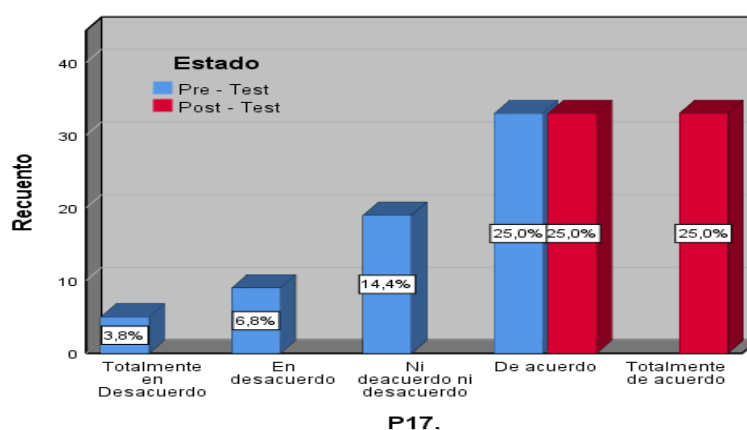


Figura 42: Resultados descriptivos sobre la pregunta 17 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Observamos que más del 50% está de acuerdo que la automatización de noticias es favorable para los lectores y esto hace que pueden encontrar noticias las

24 horas de diferentes fuentes de comunicación y pueden acceder de diferentes dispositivos.

Tabla 49  
Resultados Descriptivos sobre la pregunta 18 (Pre Test)

**P18. ¿La aplicación web scraping, ha hecho que se reduzcan los costos de publicación?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	8	12,1	12,1	12,1
	En desacuerdo	6	9,1	9,1	21,2
	Ni de acuerdo ni desacuerdo	11	16,7	16,7	37,9
	De acuerdo	41	62,1	62,1	100,0
	Total	66	100,0	100,0	

Tabla 50  
Resultados Descriptivos sobre la pregunta 18 (Post Test)

**P18. ¿La aplicación web scraping, ha hecho que se reduzcan los costos de publicación?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	35	53,0	53,0	53,0
	Totalmente de acuerdo	31	47,0	47,0	100,0
	Total	66	100,0	100,0	

**P18. ¿La aplicación web scraping, ha hecho que se reduzcan los costos de publicación? por Estado**

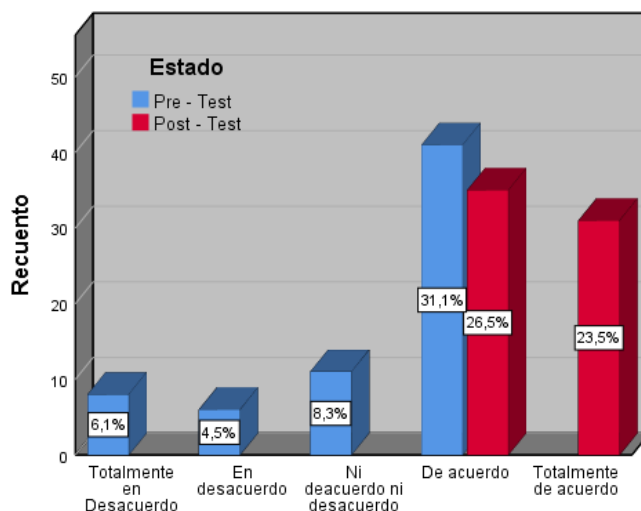


Figura 43: Resultados descriptivos sobre la pregunta 18 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Esto es parte de las mejoras significativas, ya que se de ser un trabajo manual para la recolección y publicación de noticias se pasó a la automatización de una

manera que los procesos son controlados si es que se presentara algún inconveniente. Todo lo que se redujo en costos ahora se invierte en hosting con los recursos suficientes que soporte el tráfico de las visitas y una BD de acorde con todo el contenido.

Tabla 51  
Resultados Descriptivos sobre la pregunta 19 (Pre Test)

**P19. ¿La tecnología genera progreso para ojotecnologico?**

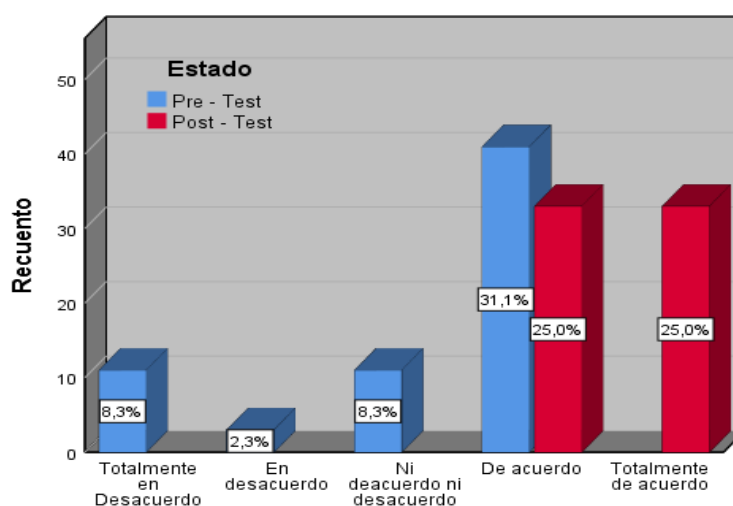
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	11	16,7	16,7	16,7
	En desacuerdo	3	4,5	4,5	21,2
	Ni de acuerdo ni desacuerdo	11	16,7	16,7	37,9
	De acuerdo	41	62,1	62,1	100,0
	Total	66	100,0	100,0	

Tabla 52  
Resultados Descriptivos sobre la pregunta 19 (Post Test)

**P19. ¿La tecnología genera progreso para ojotecnologico?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	33	50,0	50,0	50,0
	Totalmente de acuerdo	33	50,0	50,0	100,0
	Total	66	100,0	100,0	

**P19. ¿La tecnología genera progreso para ojotecnologico? por Estado**



P19.

Figura 44: Resultados descriptivos sobre la pregunta 19 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Podemos notar que el progreso tiene una significancia para el portal de noticias ojotecnologico. Y esto se debe que cada día incrementa la cantidad de visitas y seguidores, el cual motiva que siempre estemos haciendo mejores en favor de nuestros lectores.

Tabla 53  
Resultado Descriptivo sobre la pregunta 20 (Pre Test)

**P20. ¿Cree usted que se reducirá personal pro la automatización de publicación de noticias tecnológicas?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	9	13,6	13,6	13,6
	En desacuerdo	4	6,1	6,1	19,7
	Ni de acuerdo ni desacuerdo	16	24,2	24,2	43,9
	De acuerdo	37	56,1	56,1	100,0
	Total	66	100,0	100,0	

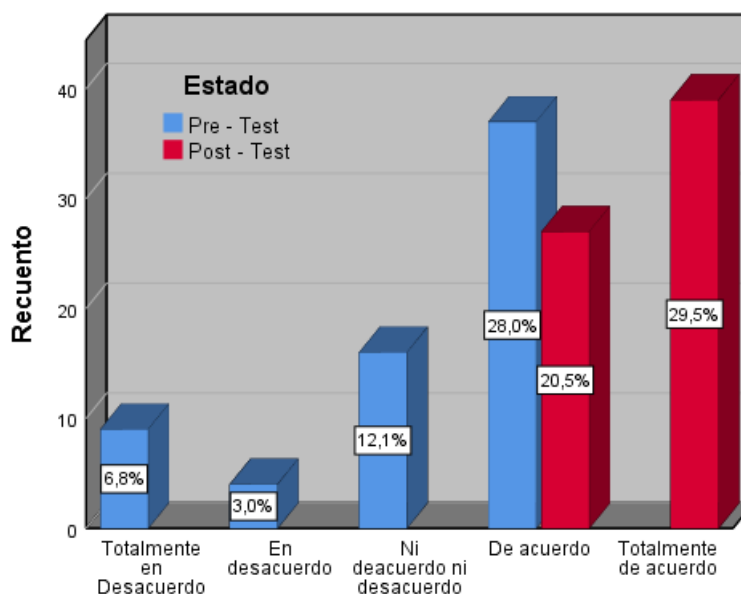
Tabla 54  
Resultado Descriptivo sobre la pregunta 20 (Post Test)

**P20. ¿Cree usted que se reducirá personal pro la automatización de publicación de noticias tecnológicas?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	27	40,9	40,9	40,9
	Totalmente de acuerdo	39	59,1	59,1	100,0
	Total	66	100,0	100,0	



**P20. ¿Cree usted que se reducirá personal pro la automatización de publicación de noticias tecnológicas? por Estado**



**P20.**

Figura 45: Resultados descriptivos sobre la pregunta 20 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Claro está que ya no se contratara personal para la publicación de noticias, pero todo eso ahorro con lleva a invertir en un hosting con los recursos y la capacidad suficiente para que el portal web sea soportado para la gran cantidad de noticias.

### 3.1.4.3 Sobre la dimisión de Calidad de servicio

Tabla 55

Resultado Descriptivo sobre la pregunta 21 (Pre Test)

**P21. ¿El Buscador de noticias localiza fácilmente la información?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	3	4,5	4,5	4,5
	En desacuerdo	8	12,1	12,1	16,7
	Ni de acuerdo ni desacuerdo	26	39,4	39,4	56,1
	De acuerdo	29	43,9	43,9	100,0
	Total	66	100,0	100,0	

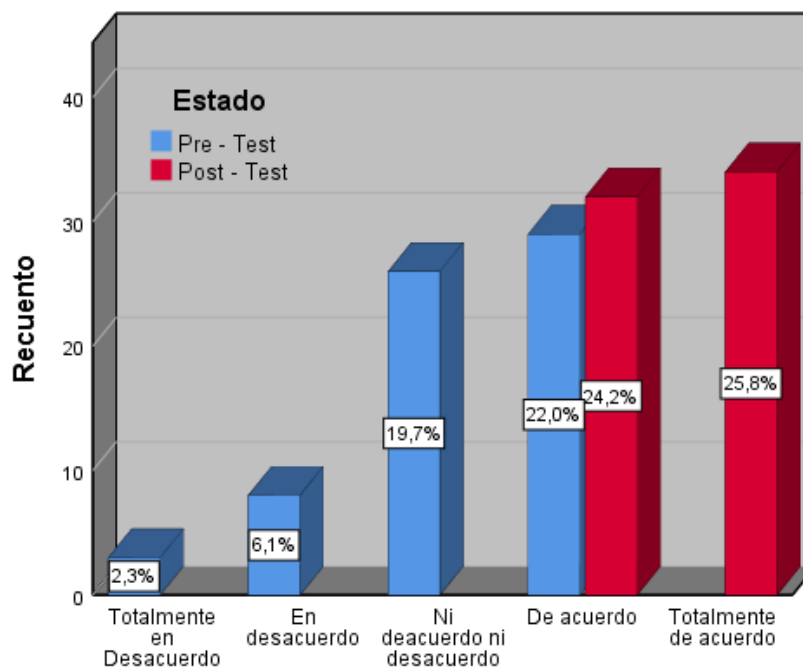
Tabla 56

Resultado Descriptivo sobre la pregunta 21 (Post Test)

**P21. ¿El Buscador de noticias localiza fácilmente la información?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	32	48,5	48,5	48,5
	Totalmente de acuerdo	34	51,5	51,5	100,0
	Total	66	100,0	100,0	

**P21. ¿El Buscador de noticias localiza fácilmente la información? por Estado**



**P21.**

Figura 46: Resultados descriptivos sobre la pregunta 21 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Podemos observar que los lectores que realizan búsquedas dentro de nuestro portal web están satisfechos con los resultados ya que pueden encontrar lo que les interesa aprender o estar enterados.

Tabla 57  
Resultado Descriptivo sobre la pregunta 22 (Pre Test)

**P22. ¿La actualización de noticias en nuestra web es constante?**

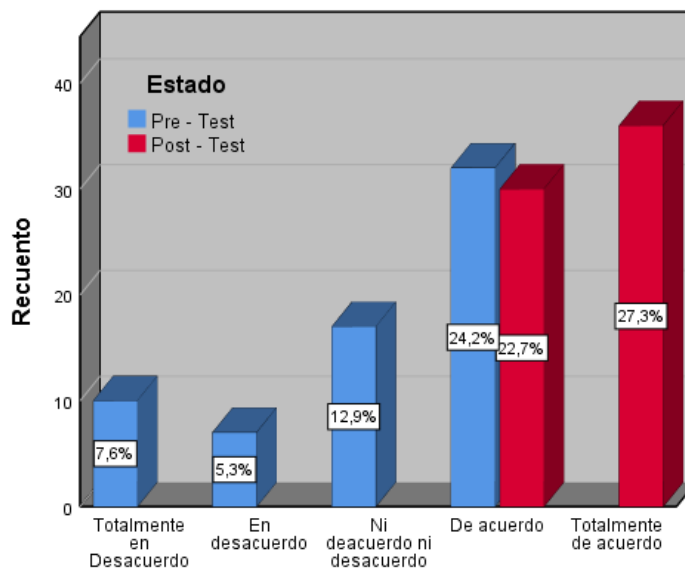
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	10	15,2	15,2	15,2
	En desacuerdo	7	10,6	10,6	25,8
	Ni de acuerdo ni desacuerdo	17	25,8	25,8	51,5
	De acuerdo	32	48,5	48,5	100,0
	Total	66	100,0	100,0	

Tabla 58  
Resultado Descriptivo sobre la pregunta 22 (Post Test)

**P22. ¿La actualización de noticias en nuestra web es constante?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	30	45,5	45,5	45,5
	Totalmente de acuerdo	36	54,5	54,5	100,0
	Total	66	100,0	100,0	

**P22. ¿La actualización de noticias en nuestra web es constante? por Estado**



**P22.**

Figura 47: Resultados descriptivos sobre la pregunta 22 (Pre y Post)

Fuente: Base de datos de la encuesta

**Interpretación:** Podemos observar que los lectores están conformes porque cada vez que ingresan al portal de noticias ojotecnologico encuentra variedad de artículos relevantes sobre actualidad tecnológica.

Tabla 59  
Resultados Descriptivos sobre la pregunta 23 (Pre Test)

**P23. ¿Desde su punto de vista existe facilidad de acceso a la información en el portal ojotecnologico?**

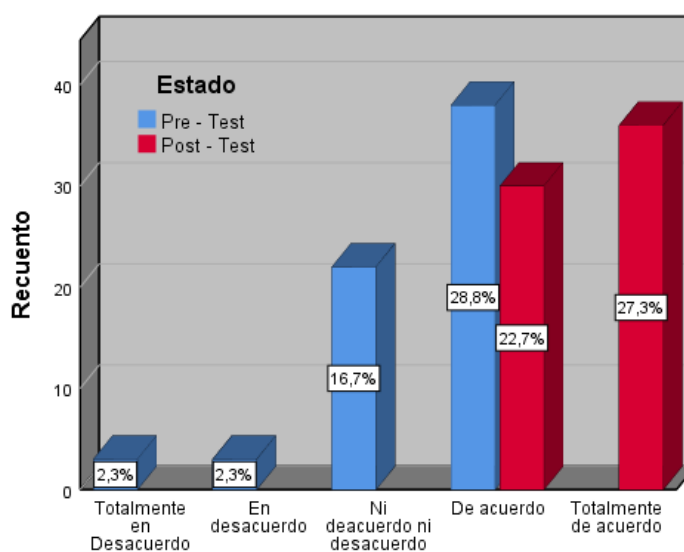
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	3	4,5	4,5	4,5
	En desacuerdo	3	4,5	4,5	9,1
	Ni de acuerdo ni desacuerdo	22	33,3	33,3	42,4
	De acuerdo	38	57,6	57,6	100,0
	Total	66	100,0	100,0	

Tabla 60  
Resultado Descriptivo sobre la pregunta 23 (Post Test)

**P23. ¿Desde su punto de vista existe facilidad de acceso a la información en el portal ojotecnologico?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	30	45,5	45,5	45,5
	Totalmente de acuerdo	36	54,5	54,5	100,0
	Total	66	100,0	100,0	

**P23. ¿Desde su punto de vista existe facilidad de acceso a la información en el portal ojotecnologico? por Estado**



**P23.**

Figura 48: Resultados descriptivos sobre la pregunta 23 (Pre y Post)  
Fuente: Base de datos de la encuesta

**Interpretación:** Con estos resultados podemos decir que el portal de ojotecnología es simple para que los usuarios puedan encontrar la noticia que están buscando y pueden enterarse de las últimas novedades en los avances de la tecnológica que cada día crece de manera ascendente.

Tabla 61  
Resultado Descriptivo sobre la pregunta 24 (Pre Test)

**P24. ¿El tiempo de carga de la web es lo que usted esperaba?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Totalmente en Desacuerdo	10	15,2	15,2	15,2
	En desacuerdo	8	12,1	12,1	27,3
	Ni de acuerdo ni desacuerdo	12	18,2	18,2	45,5
	De acuerdo	36	54,5	54,5	100,0
	Total	66	100,0	100,0	

Tabla 62  
Resultado Descriptivo sobre la pregunta 24 (Post Test)

**P24. ¿El tiempo de carga de la web es lo que usted esperaba?**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	34	51,5	51,5	51,5
	Totalmente de acuerdo	32	48,5	48,5	100,0
	Total	66	100,0	100,0	

**P24. ¿El tiempo de carga de la web es lo que usted esperaba? por Estado**

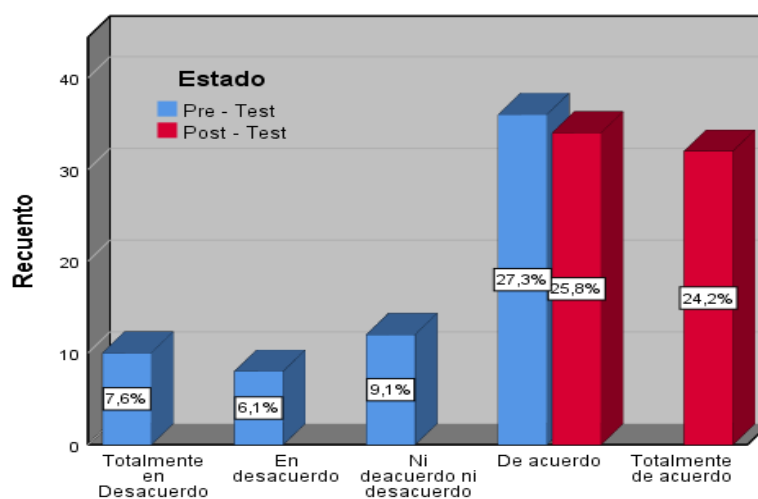


Figura 49: Resultados descriptivos sobre la pregunta 24 (Pre y Post)  
Fuente: Base de datos de la encuesta

**Interpretación:** Se trabajó bastante el tiempo de carga de nuestro portal ojotecnologico, y este punto es muy importante para que los usuarios permanezcan dentro de nuestro sitio por bastante tiempo, y luego de aplicar la web scraping, al aumentar la cantidad de noticias y trafico notamos claramente que los lectores están conformes con la velocidad de carga de la página.

### 3.2 Prueba de Normalidad

Para ello se utilizó la aplicación estadística IBM Statistics V25. El cual tiene dentro de sus herramientas la prueba de normalidad el cual utilizaremos en análisis de resultados para la investigación.

#### Nivel de significancia

- ✓  $\alpha = 0.05$  (5% de nivel de significancia) y un (95% nivel de confianza)

#### Decisión

- ✓ Si el Sig o P-valor de la muestra es  $\geq \alpha$ , entonces se acepta la Hipotesis Nula,  $H_0$  = Los datos provienen de una distribución normal-paramétrica.
- ✓ Si el Sig o P-valor de la muestra es  $< \alpha$ , entonces se acepta la hipótesis alterna,  $H_1$  = Los datos no provienen de una distribución normal.

#### Método

- ✓ Kolmogorov-Smirnov
- ✓ **Variable Independiente:** Aplicación Web Scraping (Pre y Post)

**$H_0$ :** Los datos de la muestra de la variable independiente, aplicación web scraping SI son normales.

**$H_1$ :** Los datos de la muestra de la variable independiente, aplicación web scraping No son normales.

- ✓ **Variable Dependiente:** Publicación de noticias (Pre y Post)

**H<sub>0</sub>:** Los datos de la muestra de la variable dependiente, publicación de noticias SI son normales.

**H<sub>1</sub>:** Los datos de la muestra de la variable dependiente, publicación de noticias No son normales.

A continuación, se presenta la prueba de normalidad PRE y POST TEST.

Tabla 63  
Prueba de normalidad del PRE y POST TEST de la VI y VD

		Pruebas de normalidad					
		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estado	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Web Scraping	Pre - Test	,233	66	,000	,875	66	,000
	Post - Test	,120	66	,019	,964	66	,053
Publicación de Noticias	Pre - Test	,112	66	,038	,955	66	,018
	Post - Test	,113	66	,037	,978	66	,288

a. Corrección de significación de Lilliefors

Fuente: Base de datos de la encuesta, analizado en SPSS v.25 2019

De acuerdo a los resultados obtenidos, se establece lo siguiente:

La Sig o P-Valor es menor que el nivel de significancia, entonces se acepta la Hipótesis alterna, por lo tanto la distribución de los datos, No son normales para el pre y post test.

### 3.3 Contrastación de las Hipótesis

#### 3.3.1 Contrastación de la Hipótesis General

**H.G.** La Implementación de una aplicación Web Scraping mejora significativamente la publicación de noticias tecnológicas en el portal *Web ojoTecnológico*.

**Hipótesis estadísticas para el contraste:**

**Hipótesis de la investigación (H<sub>i</sub>):** Existe influencia directa y significativa entre la aplicación web scraping y la mejora significativa de publicación de noticias tecnológicas en el portal *Web Ojotecnologico*. ((H<sub>i</sub>):  $p < \infty$ ).

**Hipótesis nula (H<sub>0</sub>):** No Existe influencia directa y significativa entre la aplicación web scraping y la mejora significativa de publicación de noticias tecnológicas en el portal *Web Ojotecnologico*. **((H<sub>0</sub>): p ≥ ∞).**

Para confirmar la hipótesis general se procedió a gestionar al paquete estadístico para ciencias sociales (SPSS Versión 25) que correlacionara ambas variables, y se obtuvieron los siguientes resultados.

Tabla 64  
Resultados de la prueba t para Web Scraping

		Prueba de muestras independientes								
		Prueba de Levene de igualdad de varianzas			prueba t para la igualdad de medias					
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
									Inferior	Superior
Web Scraping	Se asumen varianzas iguales	28,957	,000	-32,927	130	,000	-1,61869	,04916	-1,71594	-1,52143
	No se asumen varianzas iguales			-32,927	84,530	,000	-1,61869	,04916	-1,71644	-1,52094

Fuente: Información recabada de la encuesta, analizado en SPSS v.25 2019



Tabla 65  
Resultados de la prueba t para publicación de noticias

Prueba de muestras independientes										
		Prueba de Levene de igualdad de varianzas			prueba t para la igualdad de medias					
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
									Inferior	Superior
Publicación de Noticias	Se asumen varianzas iguales	47,880	,000	-30,326	130	,000	-1,55429	,05125	-1,65569	-1,45289
	No se asumen varianzas iguales			-30,326	93,172	,000	-1,55429	,05125	-1,65607	-1,45252

Fuente: Información recabada de la encuesta, analizado en SPSS v.25 2019

### Nivel de significancia

✓  $\alpha = 0.05$  (5% nivel de significancia) (95% nivel de confianza)

### Decisión

- ✓ Si  $p\text{-valor} < \alpha$  entonces se acepta la Hipótesis alterna (H1).
- ✓ Si  $p\text{-valor} \geq \alpha$  entonces se acepta la Hipótesis nula (Ho)

### Conclusión

Al mostrar un nivel de significancia bilateral de 0,000 que es mínimo al valor aceptable del indicador de decisión de 0.05 (5%), lo que demuestra la validez confiabilidad de los resultados del indicador de contraste, de tal manera se rechaza la hipótesis nula y se acepta la hipótesis de la investigación.

Con lo cual, existe una diferencia significativa en las medias de las muestras pre test y post test, por lo tanto, se concluye que:

“La Implementación de una aplicación Web Scraping mejora significativamente la publicación de noticias tecnológicas en el portal *Web ojoTecnológico*.”

### 3.3.2 Contrastación de Hipótesis Específicas

#### 3.3.2.1 Contrastación de Hipótesis Específica 1

**H.E.1** El Aumento de la cantidad de noticias publicadas por día, mejora y automatiza la publicación de noticias en la *web ojotecnológico*.

#### **Hipótesis estadísticas para el contraste:**

**Hipótesis de la investigación (Hi):** Existe influencia directa y significativa entre la cantidad de noticias publicadas y la automatización de publicación de noticias tecnológicas en el portal *Web Ojotecnologico*. ((Hi):  $p < \infty$ ).

**Hipótesis nula (Ho):** No Existe influencia directa y significativa entre la cantidad de noticias publicadas y la automatización de publicación de noticias tecnológicas en el portal *Web Ojotecnologico*. ((Hi):  $p < \infty$ ).

Parar tener certeza de la investigación se comprobó la HG en el software estadístico (SPSS Versión 25) el cual permite correlacionar las dos variables y se tiene como resultado lo siguiente.

Tabla 66  
Resultados de la prueba t para prueba de HE 1.

Prueba de muestras independientes										
		Prueba de Levene de igualdad de varianzas				prueba t para la igualdad de medias				
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
									Inferior	Superior
Cantidad de Noticias Por día	Se asumen varianzas iguales	52,760	,000	-	130	,000	-1,21970	,08109	-	-1,05926
				15,041					1,38013	
	No se asumen varianzas iguales			-	86,607	,000	-1,21970	,08109	-	-1,05850
				15,041					1,38089	

Fuente: Información recabada de la encuesta, analizado en SPSS v.25 2019

### Nivel de significancia

✓  $\alpha = 0.05$  (5% nivel de significancia) (95% nivel de confianza)

### Decisión

- ✓ Si  $p\text{-valor} < \alpha$  entonces se acepta la Hipótesis alterna (H1).
- ✓ Si  $p\text{-valor} \geq \alpha$  entonces se acepta la Hipótesis nula (Ho)

### Conclusión

Como hemos obtenido un nivel de significancia bilateral de 0,000 que es menor al valor aceptable del indicador de decisión de 0.05 (5%), lo que demuestra la validez de los resultados del indicador de contraste, de tal manera se rechaza la hipótesis nula y se acepta la hipótesis específica 1.

Con lo cual, existe una diferencia significativa en las medias de las muestras pre test y post test, por lo tanto, se concluye que:

“El Aumento de la cantidad de noticias publicadas por día, mejora y automatiza la publicación de noticias en la *web ojotecnológico*.”

### 3.3.2.2 Contratación de Hipótesis Específica 2

**H.E.2** El Incremento de la cantidad de visitas por día demuestra la afluencia de usuarios y a su vez mejora la publicación de noticias en la *web ojotecnológico*.

#### Hipótesis estadísticas para el contraste:

**Hipótesis de la investigación (Hi):** Existe influencia directa y significativa entre la cantidad de vistas y la mejora de publicación de noticias tecnológicas en el portal *Web Ojotecnologico*. ((Hi):  $p < \infty$ ).

**Hipótesis nula (Ho):** No existe influencia directa y significativa entre la cantidad de vistas y la mejora de publicación de noticias tecnológicas en el portal *Web Ojotecnologico*. ((Hi):  $p < \infty$ ).

Para tener certeza de la investigación se comprobó la HG en el software estadístico (SPSS Versión 25) el cual permite correlacionar las dos variables y se tiene como resultado lo siguiente.

Tabla 67  
Resultados de la prueba t para prueba de HE 2.

Prueba de muestras independientes										
		Prueba de Levene de igualdad de varianzas				prueba t para la igualdad de medias				
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
									Inferior	Superior
Cantidad de visitas por día	Se asumen varianzas iguales	15,079	,000	-	130	,000	-2,53030	,06539	-	-2,40094
									2,65967	

No se	-	103,574	,000	-2,53030	,06539	-	-2,40063
asumen		38,696					2,65998
varianzas							
iguales							

Fuente: Información recabada de la encuesta, analizado en SPSS v.25 2019

### Nivel de significancia

- ✓  $\alpha = 0.05$  (5% nivel de significancia) (95% nivel de confianza)

### Decisión

- ✓ Si  $p\text{-valor} < \alpha$  entonces se acepta la Hipótesis alterna (H1).
- ✓ Si  $p\text{-valor} \geq \alpha$  entonces se acepta la Hipótesis nula (Ho)

### Conclusión

Al haber obtenido un nivel de significancia bilateral de 0,000 que es menor al valor aceptable del indicador de decisión de 0.05 (5%), lo que demuestra la validez de los resultados del indicador de contraste, por lo tanto, se rechaza la hipótesis nula y se acepta y valida la hipótesis específica 2.

Con lo cual, existe una diferencia significativa en las medias de las muestras pre test y post test, por lo tanto, se concluye que:

“El Incremento de la cantidad de visitas por día demuestra la afluencia de usuarios y a su vez mejora la publicación de noticias en la *web ojotecnológico*.”

#### 3.3.2.3 Contrastación de Hipótesis Específica 3

**H.E.3** Si se garantiza el nivel de usabilidad entonces se facilita el manejo del aplicativo por parte de los usuarios y así mismo se mejora la publicación de noticias en la *web ojotecnológico*.

### Hipótesis estadísticas para el contraste:

**Hipótesis de la investigación (Hi):** Existe influencia directa y significativa entre el nivel de usabilidad y la mejora de publicación de noticias tecnológicas en el portal *Web Ojotecnologico*. ((Hi):  $p < \infty$ ).

**Hipótesis nula (Ho):** No existe influencia directa y significativa entre el nivel de usabilidad y la mejora de publicación de noticias tecnológicas en el portal *Web Ojotecnologico*. ((Hi):  $p < \infty$ ).

Para tener certeza de la investigación se comprobó la HG en el software estadístico (SPSS Versión 25) el cual permite correlacionar las dos variables y se tiene como resultado lo siguiente.

Tabla 68  
Resultados de la prueba t para prueba de HE 3.

Prueba de muestras independientes										
		Prueba de Levene de igualdad de varianzas				prueba t para la igualdad de medias				
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
									Inferior	Superior
Nivel de usabilidad	Se asumen varianzas iguales	15,489	,000	-	130	,000	-1,10606	,08135	-	-,94511
				13,595					1,26701	
	No se asumen varianzas iguales			-	90,814	,000	-1,10606	,08135	-	-,94445
				13,595					1,26767	

### Nivel de significancia

✓  $\infty = 0.05$  (5% nivel de significancia) (95% nivel de confianza)

### Decisión

- ✓ Si  $p\text{-valor} < \infty$  entonces se acepta la Hipótesis alterna (H1).
- ✓ Si  $p\text{-valor} \geq \infty$  entonces se acepta la Hipótesis nula (Ho)

## **Conclusión**

Al obtener un nivel de significancia bilateral de 0,000 que es menor al valor aceptable del indicador de decisión de 0.05 (5%), lo que demuestra la validez de los resultados del indicador de contraste, podemos concluir que se rechaza la hipótesis nula y se acepta y valida la hipótesis específica 3.

Con lo cual, existe una diferencia significativa en las medias de las muestras pre test y post test, por lo tanto, se concluye que:

“Si se garantiza el nivel de usabilidad entonces se facilita el manejo del aplicativo por parte de los usuarios y así mismo se mejora la publicación de noticias en la web ojotecnológico.”

## IV. DISCUSIÓN

De acuerdo a los resultados obtenidos en la investigación el nivel de significancia es muy alto entre la web scraping y la mejora de la publicación de noticias tecnológicas en el portal web ojotecnologico, sobra la Hipótesis general de nuestra investigación en el cual se presenta un valor de significancia bilateral de 0,000 que es menor error mínimo aceptable estadísticamente de 0,05. Estos datos nos permiten afirmar que la implementación de la técnica web scraping ayuda significativamente a poder automatizar procesos en cual faciliten el acceso a la información, que en nuestro caso serían los lectores del portal web ojotecnologico, que podrán contar con gran variedad sobre la actualizad de noticias tecnológicas.

Sobre la Hipótesis especifica 1, los resultados obtenidos de la investigación evidencian un nivel de significancia bilateral de 0,000 aceptable estadísticamente, y el valor de confiabilidad es alta y resalta el impacto del aumento de la cantidad de noticias publicadas por día para la automatización de las publicaciones en el portal *web ojotecnologico*, esto permite que todos los días se publiquen un promedio de 30 a 40 noticias por día de diferente fuentes de comunicación en el cual los usuario podrán encontrar gran variedad de noticias que podrán ser de su interés.

Sobre la Hipótesis especifica 2, los resultados obtenidos de la investigación evidencian un nivel de significancia bilateral de 0,000 aceptables estadísticamente y en el cual existe una diferencia significativa en las medias de las muestras pre test y post test. Esto nos permite afirmar que efectivamente las cantidades de visitas aumentos considerablemente y esto se puede validar en nuestro perfil de google analytics que nos permite monitorear desde que parte geográficamente son nuestros lectores.

Sobre la Hipótesis especifica 3, los resultados obtenidos de la investigación evidencian un nivel de significancia bilateral de 0,000 aceptables estadísticamente, en el cual existe diferencia en las medias los pre y post test. Esto nos permite afirmar que el nivel de usabilidad es muy bueno porque al haber más noticias la página la muestra de forma ordenado y los usuarios pueden acceder a la información sin ningún inconveniente desde cualquier dispositivo que se conecten.



## V. CONCLUSIONES

1. La Web Scraping se relaciona directa y significativamente en un nivel alto con la automatización de procesos para la publicación de noticias tecnológicas en el portal web ojotecnologico, Según los resultados obtenidos se determina (p-valor  $< \infty$ ) ( $0,000 < 0.05$ ), por lo cual demuestra que la automatización de procesos del procedimiento de obtener las noticias diariamente y que cumpla con un cierto números de criterios para poder filtrar y que sean noticias completas y respetando las fuentes originales de donde provienen las noticias ha sido un éxito.
2. La cantidad de noticias publicadas por día se relaciona directa y significativamente con la mejora de la publicación de noticias en el portal web ojotecnologico, según los resultados obtenidos se determina (p-valor  $< \infty$ ) ( $0,000 < 0.05$ ), por lo cual demuestra que hay un incremento considerable de noticias publicadas por día, en el pre test había u promedio de 5 a 8 noticias diarias luego de implementar la web scraping y hacer la prueba post test se llegó a un promedio de 30 a 40 noticias por día.
3. La cantidad de visitas por día se relaciona directa y significativamente con la afluencia y mejora la publicación de noticias en el portal web ojotecnologico, según los resultados obtenidos se determina (p-valor  $< \infty$ ) ( $0,000 < 0.05$ ), por lo cual demuestra que la cantidad de vistas es un indicador importante para la mejora de la publicaciones de noticias, y los resultados de google analytics del mes de abril del 2021 había un total de 369 visitas y en el mes de abril del 201 hay un total de 1426 visitas.
4. El nivel de usabilidad se relaciona directa y significativamente con el manejo del aplicativo web y la mejora de la publicación de noticias en el portal web ojotecnologico, según los resultados obtenidos se determina (p-valor  $< \infty$ ) ( $0,000 < 0.05$ ), por lo cual demuestra que la usabilidad del portal web, es aceptable les facilita la navegación del mismo para poder leer con normalidad y estar informados de las últimas novedades en el mundo de la tecnológica, así mismo con estos resultados afirmamos que el portal de noticias ojotecnologico e adaptable a los diferentes dispositivos que hoy en día los usuarios puedan contar.

## VI. RECOMENDACIONES

1. A la web scraping que ya está en funcionamiento en el portal web ojotecnologico, se debería agregar un módulo que se encargue de obtener las noticias que son tendencia en cada medio de comunicación previo análisis, y de esta manera ser publicadas como las más vistas o las más destacadas del día.
2. La web scraping de extracción de noticias puede convertirse en un servicio formando un API para que otras personas o medios de comunicación sobre todo en provincia puedan automatizar su proceso de recolección de noticias.
3. Los administradores del portal web ojotecnologico deben desplegar esfuerzos para agregar noticias que contengan video o foto galerías que actualmente no se muestran.
4. Los administradores del portal web ojotecnologico deben realizar en esfuerzos de contratar una agencia de marketing el cual realice un plan para hacer más conocido el portal web y así contar con más visitas a la página web, y estos sea retribuirle en ganancia de la publicidad que se muestra actualmente.
5. Los administradores del portal web ojotecnologico deben plantearse a futuro la creación de una app móvil.

## REFERENCIAS BIBLIOGRÁFICAS

### Bibliografía

- Appelmans, T. (2004). Web Globalization and WSDM Methodology of Web Design. *Vrije*. Obtenido de <https://wise.vub.ac.be/thesis/web-globalization-and-wsdm-methodology-web-design>
- Aular, Y. J., & R. Talavera Pereira. (2007). Metodologías para el desarrollo de ambientes de aprendizaje en entornos colaborativos: Una reflexión teórica. *Redalyc*, 7(1), 63-71. Obtenido de <http://www.redalyc.org/articulo.oa?id=90470108>
- Bastidas, M. J. (2009). «Desarrollo de una solución E-Comerce B2E para la gestión de. *Pontificia Universidad Católica del Ecuador Sede Ambato*. Obtenido de [http://rraae.org.ec/Record/0030\\_8716763af3b71715fea075ce9bc0c5fb/Details](http://rraae.org.ec/Record/0030_8716763af3b71715fea075ce9bc0c5fb/Details)
- BEEKMAN, G. (2005). *INTRODUCCIÓN A LA INFORMÁTICA*. (Sexta edición ed.). Madrid, España: PEARSON EDUCACIÓN, S.A. Obtenido de [https://tisgtacna.files.wordpress.com/2017/10/introduccion\\_a\\_la\\_informatica\\_beekman\\_ge.pdf](https://tisgtacna.files.wordpress.com/2017/10/introduccion_a_la_informatica_beekman_ge.pdf)
- Bembibre, C. (2013). Publicación. *Definición ABC*. Obtenido de <https://www.definicionabc.com/general/publicacion.php>
- Bermúdez, J. D. (2008). «Un Sistema de Escritura de Traductores de Escritura Vía. *International Latin American and Caribbean Conference for Engineering and Technology*, 1(1), 1-9. Obtenido de <https://arxiv.org/ftp/arxiv/papers/1205/1205.1642.pdf>
- Brinder, I. (2016). Tecnología Radial: publicaciones libres. *Radios libres*, 4-10. Obtenido de [https://radioslibres.net/wp-content/uploads/media/uploads/documentos/tutorial\\_16\\_publicaciones\\_libres.pdf](https://radioslibres.net/wp-content/uploads/media/uploads/documentos/tutorial_16_publicaciones_libres.pdf)
- Cenys, T., & Grigalis, A. (2013). Stateof-the-art web data extraction systems for online. *INFORMACIJOS MOKSLAI*.
- Chicaíza, D. F., & Paredes, F. D. (2015). Sistema de Gestión de E-Portafolios de Aprendizaje para la carrera de Ingeniería de Sistemas e Informática Utilizando UWEUML. *Universidad de las Fuerzas Armadas*. Obtenido de <http://repositorio.espe.edu.ec/xmlui/handle/21000/11365>
- Chung, W. (2012). Case studies of browsing web directories. *Managing web repositories in emerging*, 32(3):232–238. Obtenido de [https://www.researchgate.net/publication/257103282\\_Managing\\_web\\_repositories\\_in\\_emerging\\_economies\\_Case\\_studies\\_of\\_browsing\\_web\\_directories](https://www.researchgate.net/publication/257103282_Managing_web_repositories_in_emerging_economies_Case_studies_of_browsing_web_directories)
- Cobo, J. C. (2011). La utilización de la Ingeniería de Software en hipermedia. *Revista Ciencia UNEMI*, 102-117. Obtenido de <http://ojs.unemi.edu.ec/index.php/cienciaunemi/article/view/31/28>
- COMPUTING, R. (2018). Web scraping, el gran aliado de la innovación analítica. *computing*. Obtenido de <http://www.computing.es/analytics/noticias/1103589046201/web-scraping-gran-aliado-de-innovacion-analitica.1.html>
- Fiumara, E. F., & Baumgartner, R. (2014). Web data extraction, applications and techniques: A survey. *Knowledge-Based Systems*, 301,323.
- Hernández, R., Fernández, C., & Baptista, M. (2010). *Metodología de la investigación* (Quinta Edición ed.). Mexico: MCGRAW-HILL. Obtenido de

- [https://www.esup.edu.pe/descargas/dep\\_investigacion/Metodologia%20de%20la%20investigaci%C3%B3n%205ta%20Edici%C3%B3n.pdf](https://www.esup.edu.pe/descargas/dep_investigacion/Metodologia%20de%20la%20investigaci%C3%B3n%205ta%20Edici%C3%B3n.pdf)
- López B. (5 de 6 de 2018). *ciudadano2cero*. Obtenido de Qué es un Hosting Web, tipos de Alojamiento y cuál elegir: <https://www.ciudadano2cero.com/que-es-un-hosting-web-tipos-alojamiento-cual-elegir/>
- Maritn, A. R., & Martin, M. J. (2014). *Aplicaciones Web* (2 ed.). Madrid: Paraninfo. Obtenido de <https://books.google.es/books?id=43G6AwAAQBAJ>
- Martinez, M. (2004). Los grupos focales de discusión como método de investigación. *Heterotopía*(26), 59-72. Obtenido de <https://biblat.unam.mx/es/revista/heterotopia/articulo/los-grupos-focales-de-discusion-como-metodo-de-investigacion>
- Mishra, A. (2014). Benefits of combining WSDM and Scrum Framework development. *International Journal of Scientific & Technology research, III*, 328-330. Obtenido de <http://www.ijstr.org/paper-references.php?ref=IJSTR-0714-9426>
- Mitchell, R. (2015). *Web Scraping with Python*. Estados Unidos de America: O'Reilly Media. Obtenido de <https://yanfei.site/docs/dpsa/references/PyWebScrapingBook.pdf>
- Molina Ríos, J. . (2018). Comparación de metodologías en aplicaciones web. *3C Tecnología: glosas de innovación aplicadas a la pyme*, 1-19. Obtenido de <https://www.3ciencias.com/articulos/articulo/comparacion-metodologias-aplicaciones-web/>
- Ñaupas, Mejia, Novoa, & Villagomez. (2014). Metodología de la Investigación. 246.
- Oliva, P. (2009). Listas de chequeo como técnica de control. Obtenido de <http://www.minsal.gob.cl/portal/url/item/7cf9e499a55c4cc7e04001011f016c69.pdf>
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach* (Studen ed.). EEUU: Lawrence Erlbaum Associates, Inc.
- RTVE.ES. (2010). *manualdeestilo* . Obtenido de Directrices para los profesiones: <http://manualdeestilo.rtve.es>
- Sánchez, H., & Reyes, C. (2006). *Metodología y diseños en la investigación científica*. Lima: Visión Universitaria. Obtenido de [http://especializada.unsaac.edu.pe/cgi-bin/koha/opac-detail.pl?biblionumber=1094&query\\_desc=au%3A%22Reyes%20Meza%2C%20Carlos%22](http://especializada.unsaac.edu.pe/cgi-bin/koha/opac-detail.pl?biblionumber=1094&query_desc=au%3A%22Reyes%20Meza%2C%20Carlos%22)
- Valdés, D. P. (2007). ¿Qué son las bases de datos? *Mestros del web*. Obtenido de <http://www.maestrosdelweb.com/que-son-las-bases-de-datos/>
- Valle Rodriguez, A. N. (2009). *Metodologías de diseño usadas en ingeniería web, su vinculación con las NTICs*. Universidad Nacional de la Plata. Madrid: Repositorio Institucional de la UNLP. Obtenido de <http://sedici.unlp.edu.ar/handle/10915/4172>
- Bashi, F., & Antonio, C. (2012). Medición de la satisfacción del cliente y el impacto de la publicidad en unidades productivas de universidades públicas peruanas. *Universidad Nacional de Ingeniería*. Retrieved from <http://cybertesis.uni.edu.pe/handle/uni/3324>
- Goya, E. D. (n.d.). Medios de comunicación masiva, 125.
- León, A. M., Rivera, D. N., & Nariño, A. H. (2009). Relevancia de la Gestión por Procesos en la Planificación Estratégica y la Mejora Continua. *Eídos*, (2), 65–72. <https://doi.org/10.29019/eidos.v0i2.62>

- Murillo, D., & Saavedra, D. (2017). Web Scraping de los Perfiles y Publicaciones de una Afiliación en Google Scholar utilizando Aplicaciones Web e implementando un Algoritmo en R. *Memorias de Congresos UTP*, 0(0), 8–15.
- Pablo Cosgaya, Natalia Pano. (2012, September 10). Recursos gráficos para ediciones periódicas. Retrieved September 24, 2018, from <http://www.oert.org/recursos-graficos-para-ediciones-periodicas/>
- Patricia Hernández Salazar. (2011). La importancia de la satisfacción del usuario. *Documentación de las Ciencias de la Información*, 34(0), 349–368. [https://doi.org/10.5209/rev\\_DCIN.2011.v34.36463](https://doi.org/10.5209/rev_DCIN.2011.v34.36463)
- Rafael Menéndez-Barzanallana Asensio. Historia del desarrollo de aplicaciones Web. Universidad de Murcia. (n.d.). Retrieved September 24, 2018, from <https://www.um.es/docencia/barzana/DIVULGACION/INFORMATICA/Historia-desarrollo-aplicaciones-web.html>
- Ropero Rodríguez, J. (2010). Método general de extracción de información basado en el uso de lógica borrosa. Aplicación en portales web. Retrieved from <https://idus.us.es/xmlui/handle/11441/15542>
- Vilca, P., & Alexis, W. (2015). Sistema de extracción de información basado en ontologías para comentarios de un foro de discusión en línea en el dominio de cursos brindados por una entidad de educación superior. *Pontificia Universidad Católica del Perú*. Retrieved from <http://tesis.pucp.edu.pe/repositorio/handle/123456789/6064>
- Yujra, C., & Alfredo, H. (2014). *Web Scraping para la obtención de información actualizada de Internet con push notifications para smartphone* (Thesis). Retrieved from <http://repositorio.umsa.bo/handle/123456789/8405>

## ANEXOS

## Anexo 01: Matriz de Consistencia

Tabla A01.1:  
Matriz de Consistencia

Problemas General	Objetivos General	Hipótesis General	Variables Independiente	Indicador V.I.	Variables Dependiente	Indicador V.D.
¿De qué manera la implementación de una aplicación web scraping, mejorará la publicación de noticias tecnológicas en el portal web ojotecnologico?	Implementar una aplicación Web Scraping, para mejorar la publicación de noticias tecnológicas en el portal Web OjoTecnológico.	Si se implementa una aplicación Web Scraping, entonces se mejorará la publicación de noticias tecnológicas en el portal Web OjoTecnológico.	Aplicación Web Scraping	--	Publicación de noticias tecnológicas en el portal Web OjoTecnológico	--
Problemas Específico	Objetivos Específicos	Hipótesis Especificas				
¿En qué medida la cantidad de noticias publicadas por día de manera óptima permitirá la mejora de las publicaciones de noticias en la web ojotecnológico?	Aumentar la cantidad de noticias publicadas por día para mejorar y automatizar las publicaciones de noticias en la web ojotecnológico.	El Aumento de la cantidad de noticias publicadas por día mejora y automatiza la publicación de noticias en la web ojotecnológico.	Cantidad de noticias	Cantidad de noticias por día.	Satisfacción de lectores	Nivel de satisfacción
¿En qué medida la cantidad de visitas por día permitirá mostrar la afluencia de usuarios y a su vez la mejora de las publicaciones de noticias en la web ojotecnológico?	Aumentar la cantidad de visitas por día para mostrar la afluencia de usuarios y a su vez la mejora de las publicaciones de noticias en la web ojotecnológico.	El Incremento de la cantidad de visitas por día demuestra la afluencia de usuarios y a su vez mejora la publicación de noticias en la web ojotecnológico.	Cantidad de visitas	Cantidad de visitas por día.	Publicación de noticias	Reducción de costos
a) ¿En qué medida el nivel de usabilidad del aplicativo web permitirá facilitar el manejo a los usuarios y así mismo la mejora de las publicaciones de noticias en la web ojotecnológico?	Garantizar el nivel de usabilidad para facilitar el manejo del aplicativo por parte de los usuarios y así mismo la mejora de las publicaciones de noticias en la web ojotecnológico.	a) Si se garantiza el nivel de usabilidad entonces se facilita el manejo del aplicativo por parte de los usuarios y así mismo se mejora la publicación de noticias en la web ojotecnológico.	Usabilidad del aplicativo	Nivel de usabilidad	Calidad del servicio	Nivel de calidad

Elaboración propia.

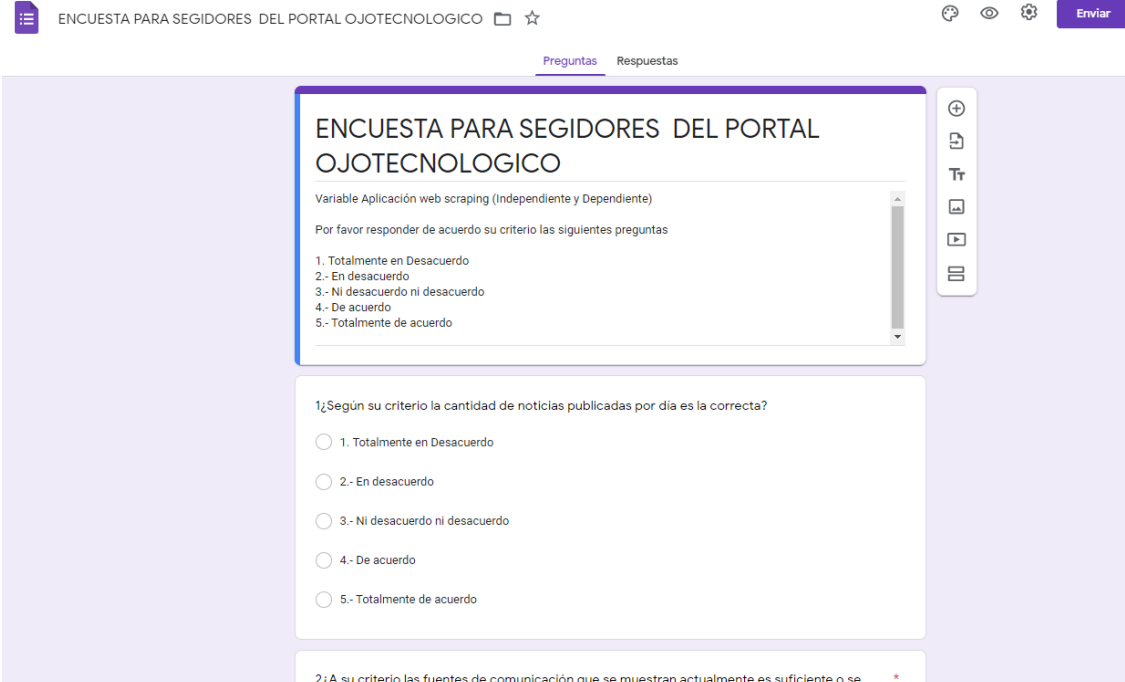
## Anexo 02: Instrumentos de recolección de datos

El instrumento de recolección de datos fue un cuestionario elaborado en la plataforma de google forms el cual genera un link público para que los encuestados pudieran ingresar y responder las preguntas

Link del formulario:

<https://forms.gle/ghXhDLNzx5npR7Ge8>

Imagen del formulario.



The image shows a Google Forms interface for a survey titled "ENCUESTA PARA SEGUIDORES DEL PORTAL OJOTECNOLOGICO". The form is in Spanish and includes the following elements:

- Title:** ENCUESTA PARA SEGUIDORES DEL PORTAL OJOTECNOLOGICO
- Variable:** Variable Aplicación web scraping (Independiente y Dependiente)
- Instruction:** Por favor responder de acuerdo su criterio las siguientes preguntas
- Legend:**
  - 1. Totalmente en Desacuerdo
  - 2.- En desacuerdo
  - 3.- Ni desacuerdo ni desacuerdo
  - 4.- De acuerdo
  - 5.- Totalmente de acuerdo
- Question 1:** 1¿Según su criterio la cantidad de noticias publicadas por día es la correcta?
  - 1. Totalmente en Desacuerdo
  - 2.- En desacuerdo
  - 3.- Ni desacuerdo ni desacuerdo
  - 4.- De acuerdo
  - 5.- Totalmente de acuerdo
- Question 2:** 2¿A su criterio las fuentes de comunicación que se muestran actualmente es suficiente o se \*

### Anexo 03: Base de datos.

#### ✓ Base de datos Pre Test

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24
2	1	4	2	1	2	1	1	2	4	2	1	1	3	1	2	4	1	1	1	2	3	4	4
3	3	3	4	1	3	1	2	4	4	3	4	3	1	3	3	1	4	1	3	3	3	4	2
4	1	1	4	3	3	1	3	3	3	4	3	2	2	1	3	3	4	4	4	3	3	4	4
3	3	2	3	1	1	2	2	4	2	2	1	3	3	2	1	4	4	4	1	3	4	3	3
3	2	3	3	3	3	3	2	4	4	4	4	1	3	3	3	4	1	1	4	4	3	3	2
3	3	3	3	3	3	3	2	3	4	4	3	3	3	3	1	3	3	4	4	4	4	3	3
4	3	4	4	3	3	2	3	4	4	4	4	1	1	1	3	3	4	3	3	3	3	4	4
3	3	3	1	3	3	2	2	1	4	3	3	1	3	2	3	4	4	4	4	4	3	4	3
3	4	4	4	2	3	2	2	1	2	1	3	3	2	2	3	3	4	4	3	4	3	4	4
3	3	4	4	2	2	3	1	3	4	4	4	1	2	1	1	4	4	4	4	3	4	4	4
2	1	4	1	3	2	2	1	3	3	4	1	2	1	1	2	3	3	4	4	3	4	3	3
4	4	4	3	2	3	1	2	4	3	4	4	2	2	3	2	3	4	4	4	3	4	3	4
3	2	2	1	1	3	3	2	3	2	1	3	2	3	3	2	2	4	4	1	1	4	2	1
2	4	3	3	3	3	1	2	3	4	3	4	2	2	2	3	3	4	3	1	4	1	4	4
1	3	3	4	3	3	3	2	4	3	4	3	2	1	1	2	2	1	3	4	4	4	4	2
4	3	4	4	2	3	3	1	4	3	3	3	1	3	3	1	4	4	4	4	3	4	4	4
4	1	4	3	3	2	2	3	4	4	4	3	1	2	1	1	1	1	4	1	3	4	3	4
3	4	4	4	3	3	1	3	4	3	4	3	3	3	1	2	4	4	4	4	4	3	4	3
4	3	3	3	3	3	2	2	3	4	4	4	1	1	3	2	2	4	2	2	1	4	1	4
4	4	3	3	1	3	2	1	4	4	4	3	2	1	2	3	4	1	4	4	2	2	3	4



3	3	4	3	2	2	1	1	4	4	4	4	1	1	2	3	4	4	4	4	4	4	4	4
4	4	3	3	2	2	2	3	4	4	3	3	3	2	1	1	3	3	4	4	3	4	4	4
3	4	3	4	1	1	1	1	4	4	3	3	3	3	2	3	3	4	3	4	4	3	3	4
4	4	3	4	1	2	1	3	3	4	3	4	2	3	1	1	4	4	4	4	4	4	3	4
4	4	4	4	1	1	2	3	3	3	4	4	2	3	2	2	4	4	4	4	4	4	4	4
4	4	4	4	2	3	2	3	4	3	4	3	2	2	3	1	4	4	4	4	4	4	4	4
4	3	4	3	1	3	1	1	3	4	4	4	3	2	3	2	4	4	4	3	3	3	4	4
3	3	3	4	1	1	3	3	3	3	3	4	1	2	2	3	3	3	4	4	4	4	4	4
4	4	4	4	1	1	1	1	4	4	4	4	2	1	2	3	4	4	3	4	4	4	4	4
4	4	3	4	1	3	1	1	4	4	4	4	1	2	1	2	4	4	4	3	4	4	4	4
3	3	4	3	2	2	2	2	3	4	4	4	3	1	1	3	4	4	4	3	2	4	1	2
4	3	4	4	3	2	2	1	4	4	4	4	1	2	3	2	2	4	1	3	4	1	4	1
4	4	3	4	3	1	2	3	4	4	3	4	2	3	3	3	3	1	4	3	4	1	4	1
4	4	3	4	2	1	2	2	4	4	3	4	2	3	1	3	3	4	4	1	1	4	4	2
4	4	4	4	3	1	2	2	3	4	4	3	1	1	3	3	3	4	4	4	3	3	3	4
4	4	4	4	1	1	1	2	4	4	4	3	2	1	1	2	3	4	4	3	4	4	4	3
3	4	4	4	2	2	1	1	4	3	4	3	3	3	1	2	4	4	3	4	4	4	4	4
3	4	3	4	2	3	3	3	3	3	4	4	3	2	1	3	3	4	4	4	3	4	3	4
4	4	3	4	1	3	1	1	3	3	4	4	1	2	2	1	4	4	3	4	4	3	3	4
3	3	4	4	1	1	1	3	4	4	4	4	3	1	1	1	4	4	4	3	3	4	4	4
4	4	4	4	3	1	3	2	4	4	3	3	1	1	2	2	4	3	4	4	4	4	4	4
4	3	4	4	2	3	2	1	4	4	4	4	2	3	2	2	2	3	3	1	4	2	4	4
4	3	4	4	2	1	2	3	4	3	3	3	1	1	3	2	2	4	2	4	3	1	3	1
4	3	3	4	1	1	1	3	3	4	3	3	2	3	2	2	2	3	4	4	3	1	4	2
4	4	3	3	2	2	1	2	4	4	3	3	2	2	2	1	1	4	1	2	3	4	4	2
3	4	4	3	2	1	2	2	4	4	4	4	2	2	2	2	3	2	4	2	3	4	1	3
4	4	4	3	1	2	3	3	4	3	4	4	2	3	1	1	4	4	4	2	3	1	3	3

4	4	3	4	2	2	3	1	4	4	4	4	3	2	2	3	2	4	1	1	3	1	4	2	
3	4	4	4	1	1	2	1	4	3	4	4	1	2	1	1	1	4	1	4	3	4	2	4	
3	4	3	4	1	2	3	2	4	3	4	3	2	2	3	2	3	3	4	3	4	3	3	4	
1	3	2	4	2	3	1	3	1	4	2	4	3	1	3	3	4	4	4	4	4	4	4	3	
2	4	2	4	3	1	2	2	4	4	4	4	3	1	3	1	4	3	3	4	3	3	3	4	
1	4	3	2	1	3	1	3	4	4	3	4	3	1	3	1	3	4	3	4	3	4	3	3	
4	4	4	4	1	2	3	1	1	4	4	4	3	1	1	1	4	4	4	3	4	3	4	3	
1	2	2	4	2	2	3	2	1	4	4	4	3	3	2	3	4	3	4	4	3	4	4	4	
4	2	4	4	2	1	2	2	2	3	2	4	3	2	1	2	4	3	3	4	4	3	3	3	
4	1	1	4	2	1	2	2	1	1	4	1	1	1	1	3	3	4	4	4	4	1	3	1	
4	4	4	4	3	2	3	1	4	3	4	2	2	2	3	2	2	1	2	4	3	1	2	4	
4	2	4	4	2	2	2	2	2	4	4	4	1	1	2	2	1	1	1	3	4	4	3	1	
2	4	2	4	1	3	1	1	2	2	4	1	1	2	1	2	4	2	4	3	3	2	3	4	
4	1	1	3	1	3	1	3	2	2	1	4	2	1	1	3	4	2	4	3	4	3	3	1	
4	4	2	3	3	3	3	2	4	3	4	4	3	1	2	1	4	4	4	1	2	2	4	4	
2	4	1	1	2	1	3	1	3	1	1	1	1	1	2	3	4	4	4	3	2	1	4	4	
1	4	2	2	2	1	1	1	3	4	4	3	1	2	1	3	4	2	1	4	2	2	4	1	
1	4	2	2	3	3	1	2	3	4	4	3	1	2	2	1	4	2	1	4	2	2	4	1	
1	4	2	2	3	3	3	3	3	3	4	4	3	1	3	3	2	4	2	1	4	2	2	4	1

## ✓ Base de datos Post Test

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24
4	5	4	4	4	4	4	5	4	5	5	4	3	3	5	3	4	4	4	4	5	5	5	4
5	5	5	4	4	4	5	4	4	5	5	4	5	5	4	4	4	4	4	4	4	5	5	5
5	4	5	4	4	4	4	5	4	4	4	5	5	4	4	4	5	4	4	5	4	4	5	5
5	5	4	4	4	4	5	5	5	5	5	5	3	3	5	4	5	5	5	4	5	5	4	5
5	4	5	5	5	4	5	5	4	5	4	4	3	5	5	5	4	4	4	5	5	5	4	4
5	5	4	4	5	4	5	4	5	4	5	5	5	4	4	4	5	5	4	5	5	5	5	4
5	5	5	4	4	5	4	4	5	5	5	4	3	4	5	5	4	5	4	5	4	4	4	5
5	5	4	4	4	5	5	4	5	4	5	5	3	4	3	3	4	5	5	4	4	5	4	4
4	4	5	5	4	5	5	5	4	4	5	5	3	5	3	5	4	4	4	4	4	4	5	5
4	5	4	4	4	4	5	4	5	5	5	5	3	5	5	3	4	4	4	5	4	5	4	5
4	4	4	5	4	5	5	4	4	4	5	5	5	3	3	4	5	5	4	5	5	4	5	4
4	4	5	4	5	4	5	4	5	5	4	5	3	5	3	5	5	4	5	5	5	4	5	5
5	4	5	4	4	5	5	5	4	5	5	4	3	3	3	5	4	4	5	5	5	5	4	5
5	5	5	4	5	4	5	5	5	4	5	5	4	4	5	5	5	4	4	4	4	5	5	5
5	4	5	5	4	4	5	5	5	4	4	4	3	3	4	5	4	4	4	5	5	5	4	5
4	5	5	5	4	5	5	5	5	4	5	5	5	5	5	4	5	4	5	5	5	5	5	5
5	5	4	5	4	5	4	5	4	5	4	4	4	3	5	5	4	5	4	4	5	5	5	5
4	4	5	4	5	4	5	5	5	5	4	5	3	3	4	3	4	5	4	4	4	5	5	4
4	4	4	4	4	5	5	4	4	4	4	4	3	3	3	4	5	4	5	4	5	4	5	5
5	4	5	5	4	5	5	4	4	4	5	4	4	4	3	5	5	5	5	5	4	4	4	4
5	4	4	5	4	5	5	4	5	4	5	4	4	4	3	5	4	5	4	5	5	5	5	4
5	4	5	5	4	5	5	5	4	4	4	4	5	3	3	3	5	4	5	4	5	4	4	4
4	4	4	4	4	5	4	5	5	4	5	4	5	5	3	5	5	4	4	4	4	5	4	5
5	4	5	4	5	5	5	4	5	4	4	4	4	4	3	5	4	4	4	4	4	4	5	4
4	4	5	5	4	4	5	5	4	4	4	5	5	3	3	5	5	5	5	5	4	5	5	4

5	4	5	5	4	4	5	5	5	5	5	5	5	5	5	4	4	5	5	5	5	4	4	4
5	4	5	4	5	4	4	4	4	5	5	4	5	4	3	3	5	5	4	5	5	4	5	4
5	4	4	5	5	4	4	4	5	5	4	5	4	4	4	5	5	4	5	5	5	5	4	5
4	5	4	4	4	5	4	4	4	5	5	5	4	5	4	5	5	4	5	4	4	4	4	5
4	5	5	4	5	4	5	4	5	5	5	4	5	4	3	4	5	4	5	5	5	5	4	5
5	5	4	4	5	5	5	5	4	4	5	5	5	4	4	3	4	5	5	4	5	4	5	4
4	5	5	4	4	4	4	4	4	4	4	5	4	4	4	4	4	5	5	4	4	5	4	5
5	5	4	4	4	4	4	5	5	5	5	4	4	4	3	3	5	5	4	4	5	5	4	4
5	4	5	4	5	4	4	4	4	5	4	4	5	4	4	3	4	4	5	4	5	4	5	5
5	4	4	4	5	5	5	5	5	5	4	5	4	5	4	5	5	4	4	5	5	5	4	4
5	4	5	5	4	4	4	4	4	5	4	5	5	4	3	4	4	5	4	4	4	4	4	5
4	5	5	5	5	4	4	4	5	5	4	5	5	4	3	3	4	5	4	5	5	4	5	4
4	5	5	5	4	5	5	4	4	5	5	4	5	5	4	3	4	4	4	4	4	4	4	5
4	5	5	5	4	4	4	5	5	4	4	4	4	4	5	3	4	5	5	5	5	5	4	4
4	4	4	5	5	4	5	4	4	5	4	4	5	4	3	3	5	4	5	4	5	4	5	5
5	4	4	5	4	5	5	4	5	4	4	5	4	3	4	4	4	4	4	5	5	5	4	4
5	5	4	5	5	5	4	5	4	5	5	5	3	3	4	5	5	5	4	5	4	4	5	4
4	4	4	4	4	5	5	4	4	4	4	5	5	4	5	4	5	4	5	5	4	5	5	4
4	4	5	5	5	5	5	5	4	4	4	5	5	4	5	5	5	5	5	5	5	4	5	4
4	5	5	5	4	4	4	5	4	5	4	4	5	5	3	3	4	4	4	5	4	5	5	5
4	5	4	4	4	5	5	5	5	4	4	4	3	3	3	5	5	5	4	5	5	4	5	4
4	4	4	4	5	5	5	5	5	4	4	4	5	5	3	5	4	4	5	5	4	4	5	5
4	4	5	4	5	5	5	5	5	5	5	5	4	5	4	4	5	5	5	5	5	4	4	5
5	4	5	4	5	4	4	4	4	5	4	4	3	3	4	3	4	4	5	4	4	4	4	5
4	4	4	4	4	5	4	4	5	5	5	4	5	4	5	3	4	4	4	5	5	4	5	5
4	5	4	4	5	4	5	5	4	5	4	4	5	5	3	5	4	5	4	4	4	4	5	5

5	4	4	5	5	5	5	5	5	5	4	5	5	3	4	5	5	4	4	5	4	4	5	4
5	4	4	5	5	5	4	5	4	4	4	4	4	4	5	4	5	4	5	5	5	5	4	5
4	5	4	5	4	5	5	4	5	4	4	4	4	3	3	4	5	4	5	5	5	4	4	5
5	5	5	4	5	5	4	5	5	5	5	5	3	4	5	5	4	5	5	5	4	5	4	4
5	5	5	5	5	4	4	5	4	4	5	4	4	5	3	3	4	5	5	5	4	5	5	4
5	5	4	5	5	4	5	4	4	5	4	5	3	4	3	5	4	4	4	5	5	5	5	4
4	4	4	4	4	4	4	4	5	5	4	4	5	5	5	5	4	5	5	5	4	4	4	4
5	4	5	5	5	4	4	5	4	4	4	5	5	3	5	4	5	5	5	4	4	4	5	5
5	4	4	5	4	4	4	5	4	5	4	4	5	5	3	3	5	4	5	5	4	5	4	4
5	5	4	4	5	5	5	4	4	5	4	4	5	5	4	5	5	5	5	5	4	5	4	4
5	5	5	5	5	4	4	5	4	5	5	4	4	5	5	3	4	5	4	4	4	5	4	5
4	5	4	5	4	4	5	4	5	4	4	4	4	5	3	3	5	4	5	4	5	5	5	4
4	4	4	4	4	5	5	5	5	5	5	5	5	4	3	5	5	5	4	5	5	4	5	5
4	5	5	4	5	5	5	5	4	5	4	4	3	4	5	5	4	4	5	4	4	5	4	4

**Anexo 04: Evidencia de similitud digital**

IMPLEMENTACIÓN DE UNA  
APLICACIÓN WEB SCRAPING Y  
LA MEJORA DE LA PUBLICACIÓN  
DE NOTICIAS TECNOLÓGICAS  
EN EL PORTAL WEB  
OJOTECNOLÓGICO

*por Chuquijajas Chuquilín Jorge Luis*

---

**Fecha de entrega:** 04-nov-2022 10:09a.m. (UTC-0500)

**Identificador de la entrega:** 1944513209

**Nombre del archivo:** Tesis\_-\_CHUQUIJAJAS\_CHUQUILIN\_JORGE\_LUIS..docx (4.25M)

**Total de palabras:** 29201

**Total de caracteres:** 142779

## IMPLEMENTACIÓN DE UNA APLICACIÓN WEB SCRAPING Y LA MEJORA DE LA PUBLICACIÓN DE NOTICIAS TECNOLÓGICAS EN EL PORTAL WEB OJOTECNOLÓGICO


### INFORME DE ORIGINALIDAD



### FUENTES PRIMARIAS

<b>1</b>	<b>docplayer.es</b> Fuente de Internet	<b>4%</b>
<b>2</b>	<b>repositorio.unfv.edu.pe</b> Fuente de Internet	<b>2%</b>
<b>3</b>	<b>repositorio.upci.edu.pe</b> Fuente de Internet	<b>2%</b>
<b>4</b>	<b>repositorio.ucv.edu.pe</b> Fuente de Internet	<b>2%</b>
<b>5</b>	<b>radioslibres.net</b> Fuente de Internet	<b>2%</b>
<b>6</b>	<b>repositorio.umsa.bo</b> Fuente de Internet	<b>2%</b>
<b>7</b>	<b>www.3ciencias.com</b> Fuente de Internet	<b>2%</b>
<b>8</b>	<b>hdl.handle.net</b> Fuente de Internet	<b>2%</b>

## Anexo 05: Autorización de publicación en repositorio



**UPCI**  
CAMINO AL ÉXITO  
UNIVERSIDAD PERUANA DE CIENCIAS E INFORMÁTICA

### FORMULARIO DE AUTORIZACIÓN PARA LA PUBLICACIÓN DE TRABAJO DE INVESTIGACIÓN O TESIS EN EL REPOSITORIO INSTITUCIONAL UPCI

**1.- DATOS DEL AUTOR**

Apellidos y Nombres: Chuguijajas Chuguilin, Jorge Luis

DNI: 46314731 Correo electrónico: jchuguijajas@gmail.com

Domicilio: Asc de Vivienda Nueva Gales 4722 415 - Cieneguilla

Teléfono fijo: \_\_\_\_\_ Teléfono celular: 941393428

**2.- IDENTIFICACIÓN DEL TRABAJO Ó TESIS**

Facultad/Escuela: Ciencias e Ingeniería

Tipo: Trabajo de Investigación Bachiller ( ) Tesis (X)

Título del Trabajo de Investigación / Tesis:  
"Implementación de una aplicación web scraping y la mejora  
de la publicación de noticias tecnológicas en el portal web  
ego tecnológico"

**3.- OBTENER:**


Bachiller ( ) Título (X) Mg. ( ) Dr. ( ) PhD. ( )

**4. AUTORIZACIÓN DE PUBLICACIÓN EN VERSIÓN ELECTRÓNICA**

Por la presente declaro que el documento indicado en el ítem 2 es de mi autoría y exclusiva titularidad, ante tal razón autorizo a la Universidad Peruana Ciencias e Informática para publicar la versión electrónica en su Repositorio Institucional (<http://repositorio.upci.edu.pe>), según lo estipulado en el Decreto Legislativo 822, Ley sobre Derecho de Autor, Art23 y Art.33.

Autorizo la publicación de mi tesis (marque con una X):  
 (X) Sí, autorizo el depósito y publicación total.  
 ( ) No, autorizo el depósito ni su publicación.

Como constancia firmo el presente documento en la ciudad de Lima, a los  
18 días del mes de noviembre de 2022.

  
 \_\_\_\_\_  
 Firma

