

UNIVERSIDAD NACIONAL DE HUANCAMELICA

(Creada por Ley N° 25265)

FACULTAD DE INGENIERÍA ELECTRÓNICA - SISTEMAS

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



TESIS

**“MODELO DE SISTEMA BASADO EN WEB SCRAPING EN LA
GESTIÓN DE INFORMACIÓN EN LA UNIDAD DE SEGUROS
PÚBLICOS Y PRIVADOS DE LA UNIDAD EJECUTORA 403 RED DE
SALUD TAYACAJA 2019”**

LÍNEA DE INVESTIGACIÓN

GESTIÓN DE TECNOLOGÍAS DE LA INFORMACIÓN

PRESENTADO POR:

Bach. ROMANI PAREJAS, VICTOR ARMANDO

Bach. SILVA MONTAÑEZ, LUIS ALBERTO

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO DE SISTEMAS

HUANCAMELICA, PERÚ

2021



UNIVERSIDAD NACIONAL DE HUANCAVELICA
(Creada por Ley N° 25265)
FACULTAD DE INGENIERÍA ELECTRÓNICA – SISTEMAS



ACTA DE SUSTENTACIÓN DE PROYECTO DE INVESTIGACIÓN VIRTUAL

Mediante el aplicativo Google Meet con enlace: meet.google.com/wuk-qmzb-goe; Unirse por (US) +1 515-518-4805 PIN: 446207179, habilitado por Secretaría Docente de la Facultad de Ingeniería Electrónica – Sistemas, en mérito a la **Resolución de Consejo de Facultad N° 224-2021-FIES-UNH** de fecha 30 de septiembre de 2021, a los 14 días del mes de octubre del año 2021, a las 11:00 A.M, se reunieron; el Jurado Calificador, conformado de la siguiente manera:

Presidente: Dr. Rafael Wilfredo ROJAS BUJAICO
Secretario: Mg. Víctor Raúl, RODRÍGUEZ PEÑA
Vocal: Dr. Fernando Viterbo SINCHE CRISPIN

Designados con Resolución N° 060-2019-DFIES-UNH, de fecha 30 de diciembre de 2019 del proyecto de investigación, Titulado:

“MODELO DE SISTEMA BASADO EN WEB SCRAPING EN LA GESTIÓN DE INFORMACIÓN EN LA UNIDAD DE SEGUROS PÚBLICOS Y PRIVADOS DE LA UNIDAD EJECUTORA 403 RED DE SALUD TAYACAJA 2019”

Cuyos autores son los graduados: Bachilleres:

Víctor Armando, ROMANÍ PAREJAS
Luis Alberto, SILVA MONTAÑEZ

A fin de proceder con la evaluación y calificación de la sustentación del proyecto de investigación, antes citado. Se dio inicio a la sustentación del proyecto de investigación en mención, a horas 11 con 05 minutos, concluyendo a horas 12 con 25 minutos.

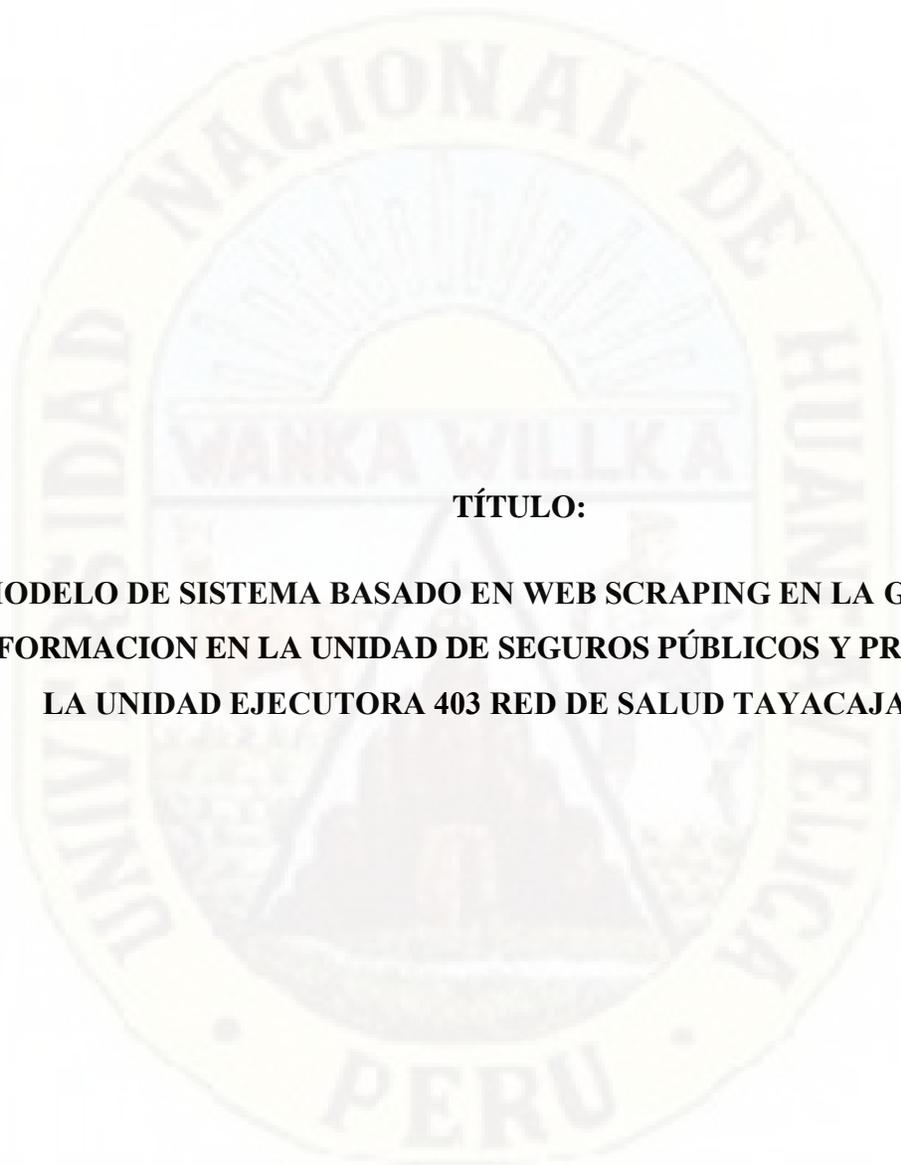
Finalizado la sustentación; se invitó al público presente y a los sustentantes a abandonar la sala de actos; y, luego de una amplia deliberación y calificación por parte del jurado, se llegó al siguiente resultado:

APROBADO POR: MAYORÍA


Dr. Rafael Wilfredo ROJAS BUJAICO
Presidente

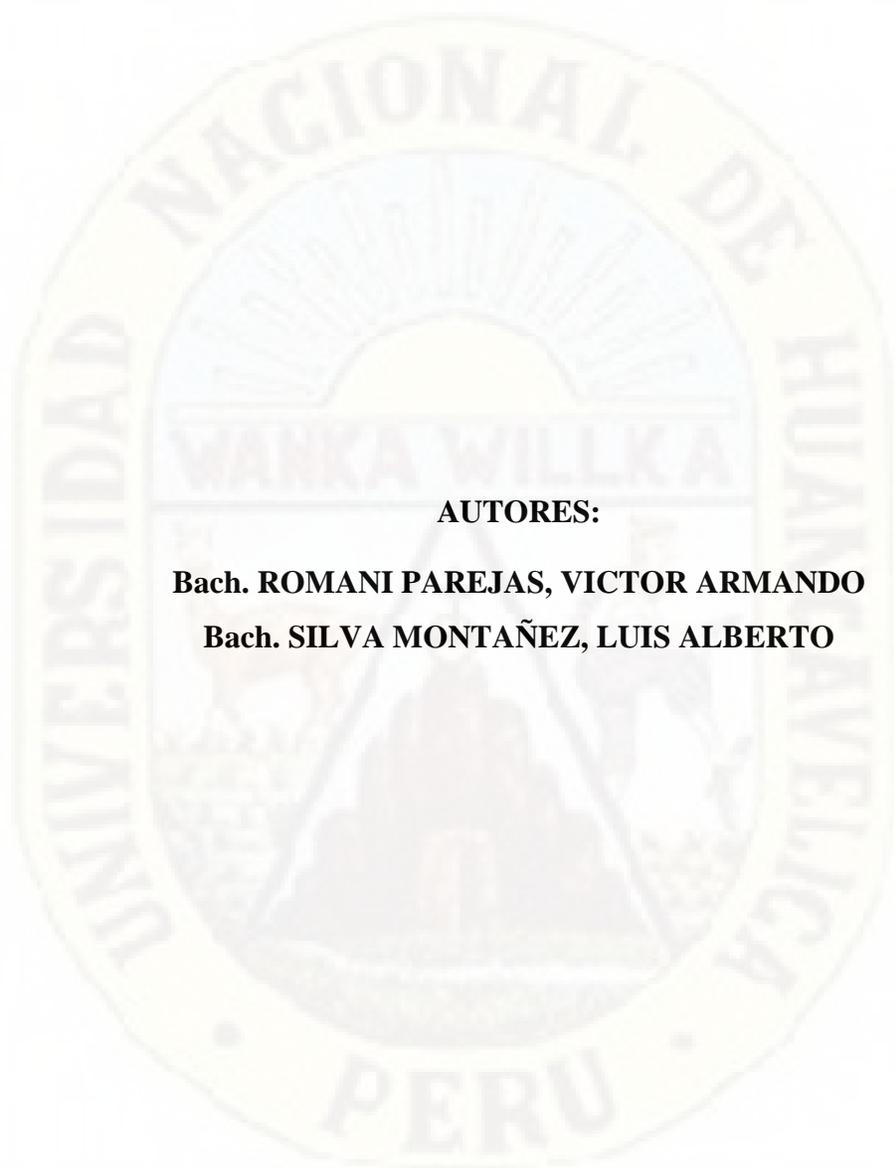

Mg. Víctor Raúl, RODRÍGUEZ PEÑA
Secretario


Dr. Fernando Viterbo SINCHE CRISPIN
Vocal



TÍTULO:

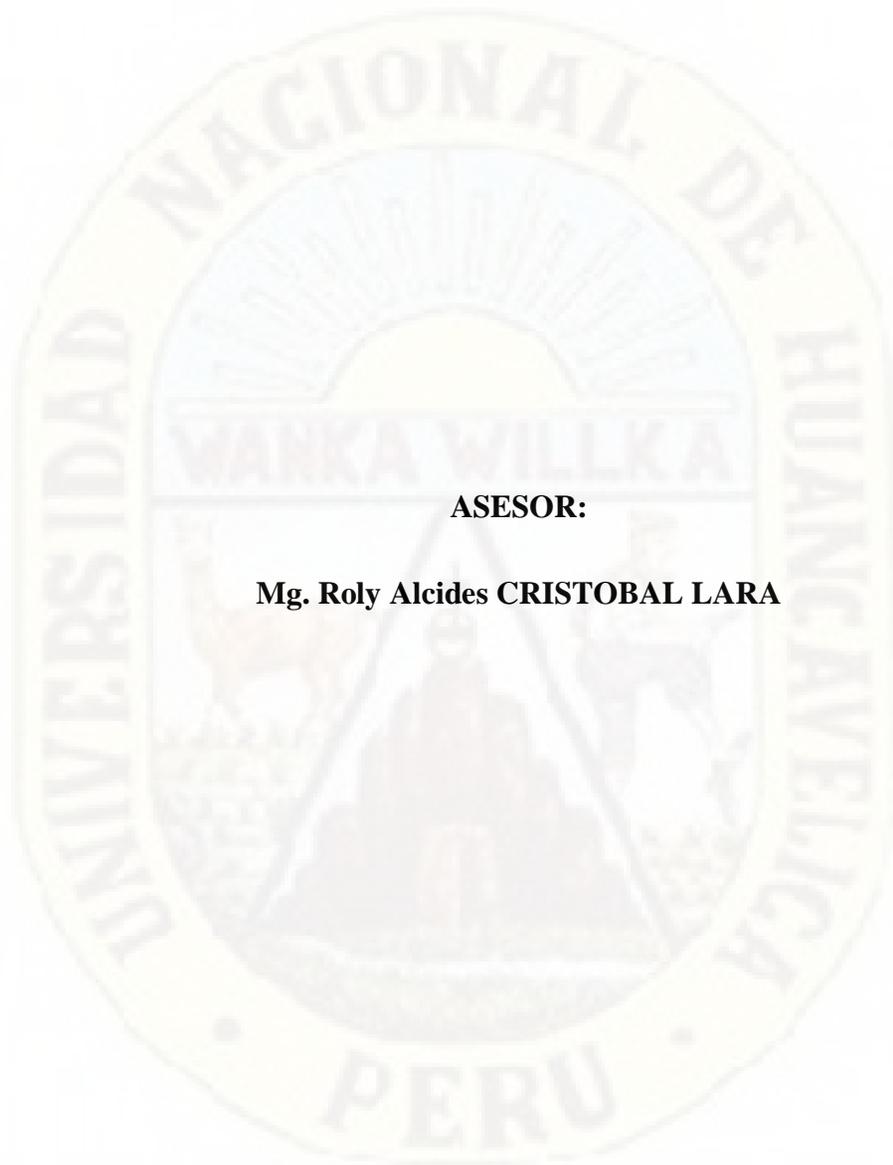
“MODELO DE SISTEMA BASADO EN WEB SCRAPING EN LA GESTIÓN DE INFORMACION EN LA UNIDAD DE SEGUROS PÚBLICOS Y PRIVADOS DE LA UNIDAD EJECUTORA 403 RED DE SALUD TAYACAJA 2019”



AUTORES:

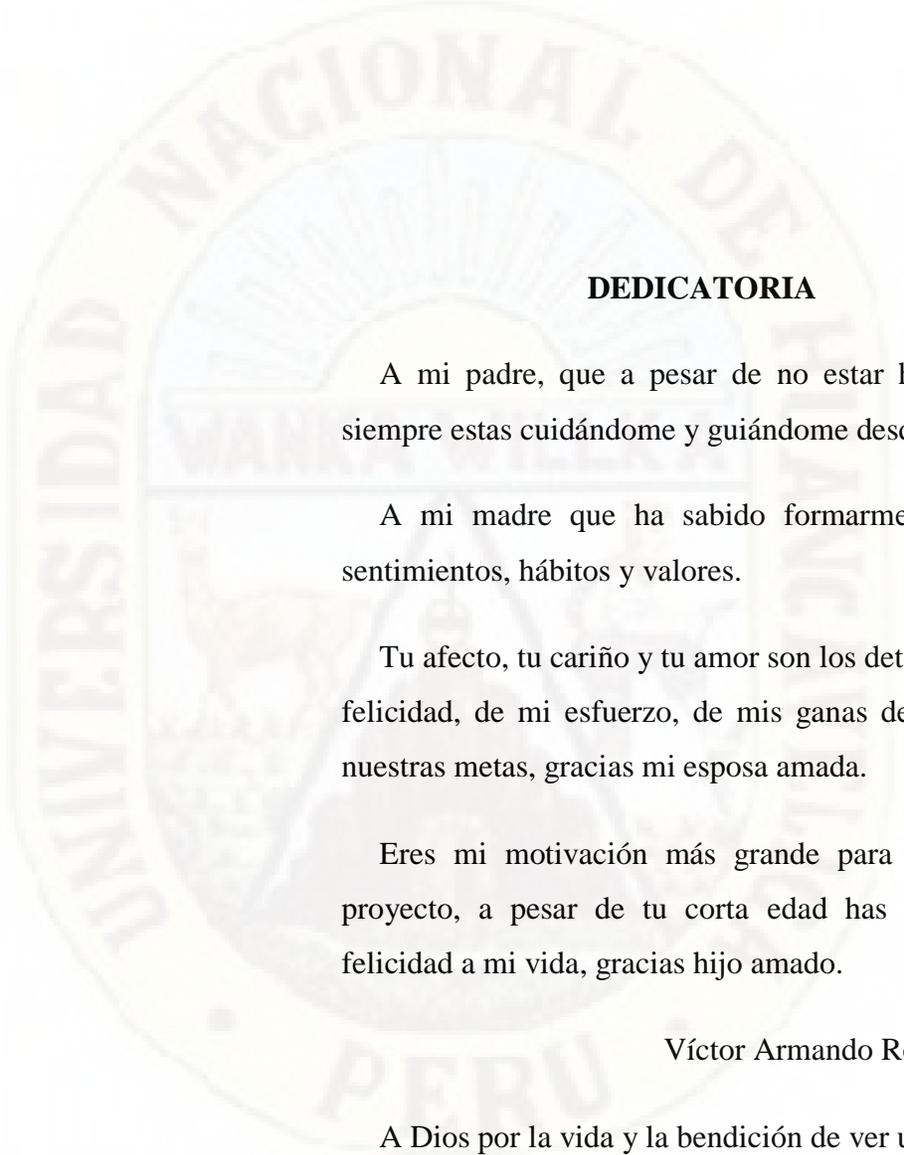
Bach. ROMANI PAREJAS, VICTOR ARMANDO

Bach. SILVA MONTAÑEZ, LUIS ALBERTO



ASESOR:

Mg. Roly Alcides CRISTOBAL LARA



DEDICATORIA

A mi padre, que a pesar de no estar hoy conmigo, siempre estas cuidándome y guiándome desde el cielo.

A mi madre que ha sabido formarme con buenos sentimientos, hábitos y valores.

Tu afecto, tu cariño y tu amor son los detonantes de mi felicidad, de mi esfuerzo, de mis ganas de lograr todas nuestras metas, gracias mi esposa amada.

Eres mi motivación más grande para concluir este proyecto, a pesar de tu corta edad has traído mucha felicidad a mi vida, gracias hijo amado.

Víctor Armando Romaní Parejas

A Dios por la vida y la bendición de ver un nuevo día.

A mis padres, por ser la fuerza y el apoyo en todo momento y ser los responsables de cultivar humanidad en mi vida.

Luis Alberto Silva Montañez

AGRADECIMIENTO

A nuestros padres por habernos apoyado incondicionalmente, por ser los principales promotores de nuestros sueños, por confiar y creer en nuestras expectativas, por los consejos, valores y principios que nos han inculcado.

A nuestros docentes de la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional de Huancavelica, por haber compartido sus conocimientos a lo largo de la preparación de nuestra profesión.

A nuestro asesor de tesis Mg. Ing. Roly Alcides Cristóbal Lara, por habernos guiado en la realización de este proyecto de tesis.

A los trabajadores de la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja, por habernos brindado la información necesaria para la realización de este proyecto de tesis.

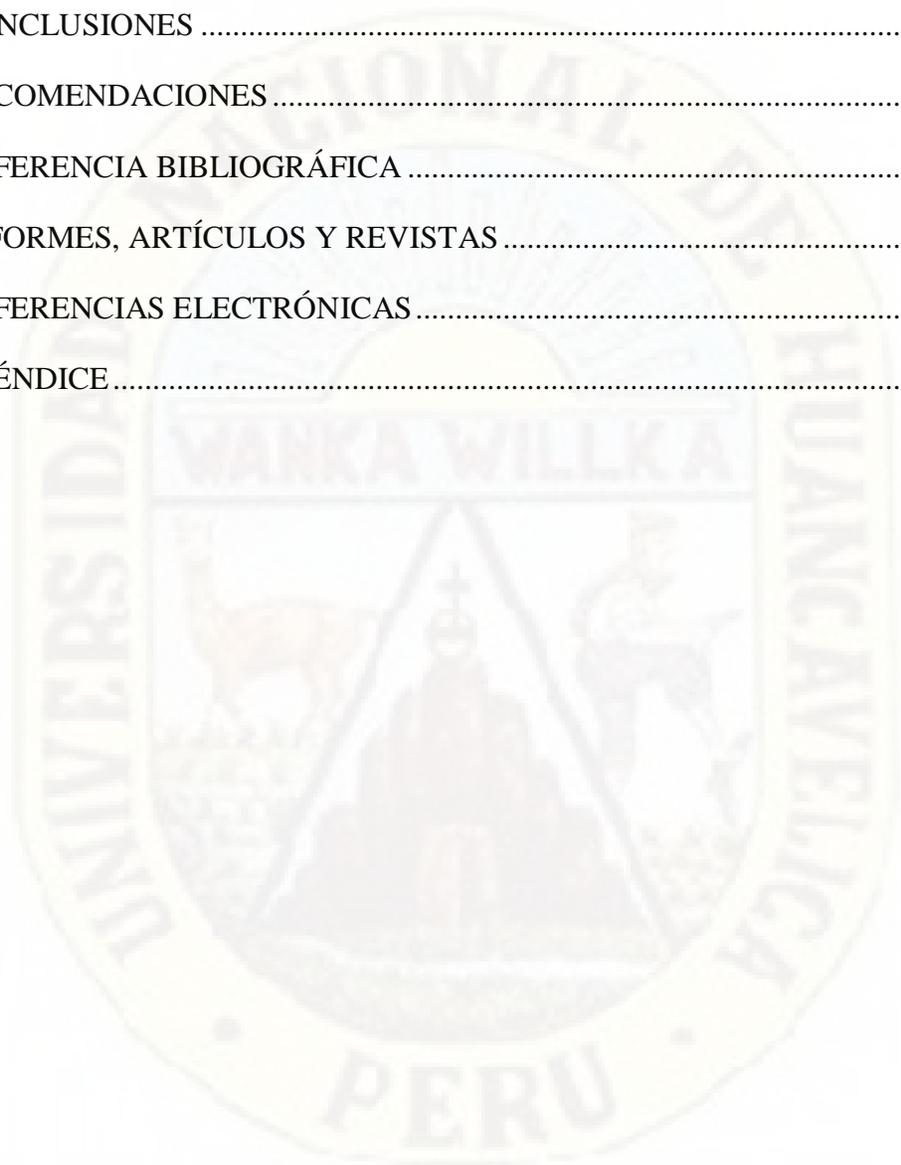
ÍNDICE

ACTA DE SUSTENTACIÓN.....	ii
TÍTULO:.....	iii
AUTORES:.....	iv
ASESOR:.....	v
DEDICATORIA.....	vi
AGRADECIMIENTO.....	vii
ÍNDICE.....	viii
ÍNDICE DE FIGURAS.....	xii
ÍNDICE DE TABLAS.....	xv
RESUMEN.....	xvii
ABSTRACT.....	xviii
INTRODUCCIÓN.....	xix
CAPÍTULO I.....	20
PLANTEAMIENTO DEL PROBLEMA.....	20
1.1. DESCRIPCIÓN DEL PROBLEMA.....	20
1.2. FORMULACIÓN DEL PROBLEMA.....	26
1.2.1. Problema General.....	26
1.2.2. Problemas Específicos.....	26
1.3. OBJETIVOS.....	26
1.3.1. Objetivo General.....	26
1.3.2. Objetivos Específicos.....	26
1.4. JUSTIFICACIÓN.....	27
1.4.1. Justificación Teórica.....	27
1.4.2. Justificación Metodológica.....	28
1.4.3. Justificación Práctica.....	28
CAPÍTULO II.....	29
MARCO TEÓRICO.....	29

2.1.	ANTECEDENTES	29
2.1.1.	Antecedentes Internacionales	29
2.1.2.	Antecedentes Nacionales	30
2.2.	BASES TEÓRICAS SOBRE EL TEMA DE INVESTIGACIÓN	33
2.2.1.	Gestión de Información	34
2.2.2.	Web Scraping.....	37
2.2.3.	ISO 25000	48
2.3.	BASES CONCEPTUALES	48
2.3.1.	Oficina Unidad de Seguros Públicos y Privados	48
2.3.2.	Web Scraping: ¿Qué programas puedo usar para ser un web scraper?	50
2.3.3.	Información.....	52
2.3.4.	Adquisición de Información	53
2.3.5.	Distribución de Información.....	53
2.3.6.	Toma de Decisiones.....	54
2.4.	DEFINICIÓN DE TÉRMINOS	54
2.5.	HIPÓTESIS	55
2.5.1.	Hipótesis General.....	55
2.5.2.	Hipótesis Específicas	55
2.6.	VARIABLES.....	56
2.6.1.	De la Hipótesis General	56
2.6.2.	Variable Dependiente:	56
2.6.3.	Variable Independiente:	56
2.7.	OPERACIONALIZACIÓN DE VARIABLES	56
CAPÍTULO III.....		58
MATERIALES Y MÉTODOS		58
3.1.	ÁMBITO TEMPORAL Y ESPACIAL.....	58
3.1.1.	Ámbito Temporal.....	58
3.1.2.	Ámbito Espacial.....	58

3.2. TIPO DE INVESTIGACIÓN	58
3.3. NIVEL DE INVESTIGACIÓN	59
3.3.1. Diseño de Investigación.....	59
3.4. POBLACIÓN, MUESTRA, MUESTREO	60
3.4.1. Población	60
3.4.2. Muestra	60
3.5. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS.....	61
3.5.1. Técnicas	61
3.5.2. Instrumentos.....	62
3.6. TÉCNICAS Y PROCESAMIENTO DE ANÁLISIS DE DATOS	62
CAPÍTULO IV.....	64
DISCUSIÓN DE RESULTADOS	64
4.1. RESULTADOS	64
4.1.1. Análisis de Información de la Variable Independiente Modelo de Sistema basado en Web Scraping.....	64
4.1.2. Análisis de Información de la Variable Dependiente Gestión de Información	68
4.1.1. Dimensión: Adquisición de Información.....	68
4.1.2. Dimensión: Distribución de Información	71
4.1.3. Dimensión: Toma de Decisiones	74
4.2. PRUEBA DE HIPÓTESIS.....	82
4.2.1. Prueba de la Hipótesis Especifica N° 01	82
4.2.2. Prueba de la Hipótesis Especifica N° 02	87
4.2.3. Prueba de la Hipótesis Especifica N° 03	91
4.3. DISCUSIÓN DE RESULTADOS	95
4.3.1. En relación a la variable independiente modelo de sistema basado en Web Scraping	95
4.3.2. De acuerdo con la dimensión Adquisición de Información.....	95

4.3.3. De acuerdo con la dimensión Distribución de Información	96
4.3.4. De acuerdo con la dimensión Toma de Decisiones	96
CONCLUSIONES	98
RECOMENDACIONES.....	100
REFERENCIA BIBLIOGRÁFICA	101
INFORMES, ARTÍCULOS Y REVISTAS	104
REFERENCIAS ELECTRÓNICAS.....	105
APÉNDICE.....	106

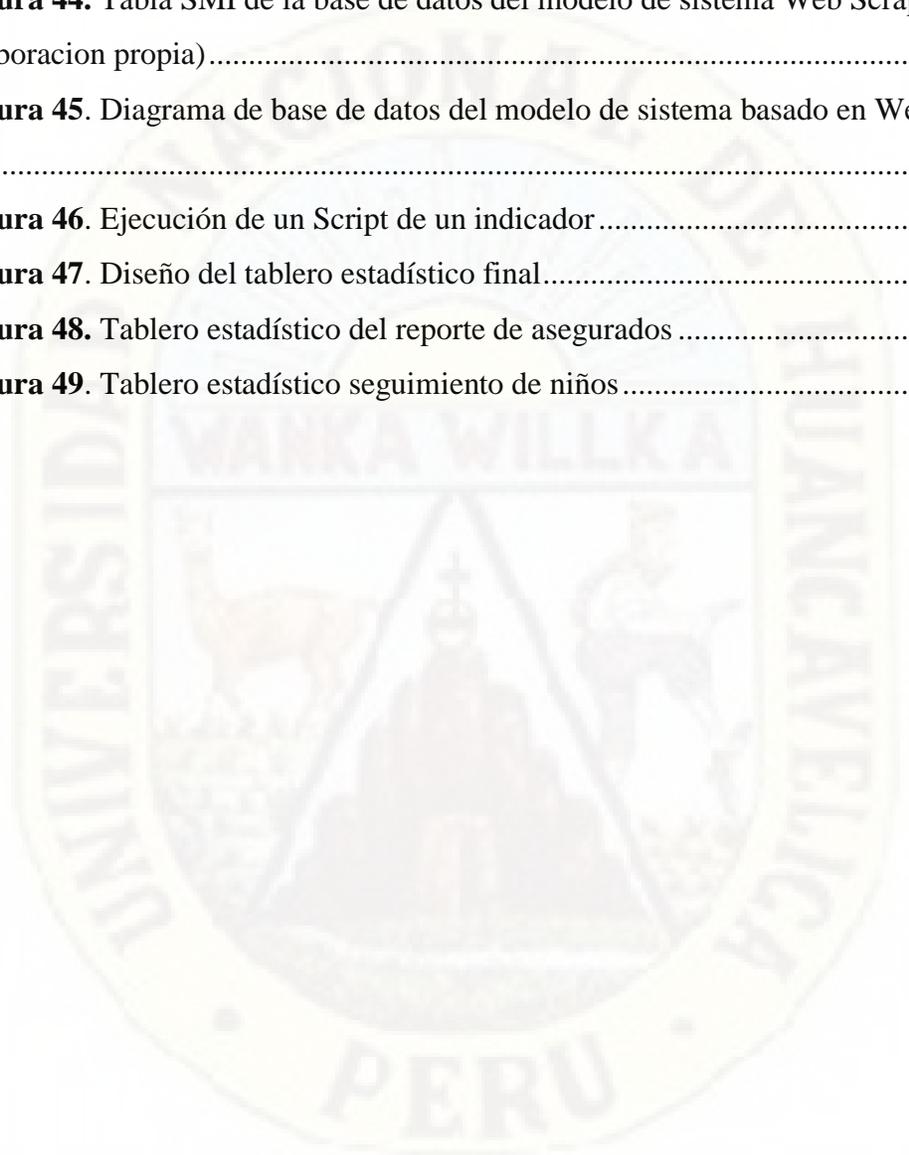


ÍNDICE DE FIGURAS

Figura 1. Resultados de la pregunta 01 del PRE TEST	23
Figura 2. Resultados de la pregunta 02 del PRE TEST	24
Figura 3. Resultados de la pregunta 03 del PRE TEST	24
Figura 4. Resultados de la pregunta 04 del PRE TEST	24
Figura 5. Resultados de la pregunta 05 del PRE TEST	25
Figura 6. El ciclo de la gestión de la información	36
Figura 7. Organigrama de la Oficina de Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja	50
Figura 8. Diseño de pre prueba y post prueba con un solo grupo	60
Figura 9. Grafica de la calidad de uso del modelo de sistema basado en Web Scraping	68
Figura 10. Resultado total de respuestas y promedio del tiempo de adquisición de información	71
Figura 11. Resultado total de respuestas y promedio del tiempo de distribución de información	74
Figura 12. Escala de Likert	74
Figura 13. Resultado de respuestas y promedio de la toma de decisiones.....	79
Figura 14. Resultado total de respuestas y promedio de condiciones para la toma de decisiones	80
Figura 15. Resultado total de respuestas y promedio del modelo de toma de decisiones	81
Figura 16. Resultado total de respuesta y promedio de estilos de toma de decisiones ...	82
Figura 17. Modelo de sistema basado en Web Scraping	113
Figura 18. Configuración de extensiones.....	114
Figura 19. Ejemplo de selector de enlaces.....	116
Figura 20. Vista previa de los selectores de enlaces	117
Figura 21. Ejemplo del grafico selector	117
Figura 22. Configuración de selector de enlaces	120

Figura 23. Selección de múltiples selectores de enlaces.....	120
Figura 24. Selectores de texto	121
Figura 25. Exportación de data as CSV	124
Figura 26. Exportación automatizada de datos	125
Figura 27. Conjunto de datos	126
Figura 28. Registro de información JSON.....	126
Figura 29. Registro de información CSV.....	127
Figura 30. Pantalla Excel	128
Figura 31. Configuración de ajustes de importación	128
Figura 32. Codificación del Web Scraping	129
Figura 33. Extracción de datos.....	154
Figura 34. Exportación de datos	155
Figura 35. Tabla atenciones de la base de datos del modelo de sistema Web Scraping (Fuente: Elaboracion propia).....	156
Figura 36. abla atenciones de la base de datos del modelo de sistema Web Scraping (Fuente: Elaboracion propia).....	156
Figura 37. abla atenciones de la base de datos del modelo de sistema Web Scraping (Fuente: Elaboracion propia).....	157
Figura 38. abla atenciones de la base de datos del modelo de sistema Web Scraping (Fuente: Elaboracion propia).....	157
Figura 39. Tabla atenciones de la base de datos del modelo de sistema Web Scraping (Fuente: Elaboracion propia).....	158
Figura 40. Tabla insumos de la base de datos del modelo de sistema Web Scraping (Fuente: Elaboracion propia).....	158
Figura 41. Tabla diagnosticos de la base de datos del modelo de sistema Web Scraping (Fuente: Elaboracion propia).....	159
Figura 42. Tabla medicamentos de la base de datos del modelo de sistema Web Scraping (Fuente: Elaboracion propia)	159

Figura 43. Tabla procedimientos de la base de datos del modelo de sistema Web Scraping (Fuente: Elaboracion propia)	160
Figura 44. Tabla SMI de la base de datos del modelo de sistema Web Scraping (Fuente: Elaboracion propia).....	160
Figura 45. Diagrama de base de datos del modelo de sistema basado en Web Scraping	161
Figura 46. Ejecución de un Script de un indicador	162
Figura 47. Diseño del tablero estadístico final.....	162
Figura 48. Tablero estadístico del reporte de asegurados	163
Figura 49. Tablero estadístico seguimiento de niños	163



ÍNDICE DE TABLAS

Tabla 1. Población Atendida en un Servicio en Salud por la Micro Red en la USPP de la UERST – Huancavelica, Año 2018	21
Tabla 2. Formatos Únicos de Atención (FUA) que no superaron las Reglas de Consistencia del Periodo 2019 Publicados por el SIS	21
Tabla 3. Definición operacional de las variables	56
Tabla 4. Resultados obtenidos en la encuesta después de la Implementación del Modelo de Sistema basado en Web Scraping Aplicando la Norma ISO/IEC 25010.	65
Tabla 5. Resumen del Modelo de la Lista de Cotejo para el Modelo de Sistema basado en Web Scraping.	66
Tabla 6. Resultados de la encuesta realizada antes y después de la Implementación del Modelo de Sistema basado en Web Scraping en la Adquisición de Información.	68
Tabla 7. Resultados de frecuencias obtenidas antes y después de la Implementación del Modelo de Sistema basado en Web Scraping en la Adquisición de Información.	69
Tabla 8. Resultados de la encuesta realizada antes y después de la Implementación del Modelo de Sistema basado en Web Scraping en la Distribución de Información.	71
Tabla 9. Resultados de frecuencias obtenidas antes y después de la Implementación del Modelo de Sistema basado en Web Scraping en la Distribución de Información.	72
Tabla 10. Resultados de la encuesta realizada antes y después de la implementación del modelo de sistema basado en Web Scraping en la toma de decisiones.	74
Tabla 11. Resultados de frecuencias obtenidas antes y después de la implementación del modelo de sistema basado en Web Scraping en la toma de decisiones.	76
Tabla 12. Pruebas de normalidad Kolmogorov-Smirnov - Shapiro-Wilk	83
Tabla 13. Normalidad de Pvalor	84
Tabla 14. Estadísticas de muestras emparejadas	84
Tabla 15. Correlaciones de muestras emparejadas.....	85
Tabla 16. Prueba de muestras emparejadas.....	85
Tabla 17. Pruebas de normalidad Kolmogorov-Smirnov - Shapiro-Wilk	88
Tabla 18. Normalidad de Pvalor	88

Tabla 19. Estadísticas de muestras emparejadas.....	89
Tabla 20. Correlaciones de muestras emparejadas.....	89
Tabla 21. Prueba de muestras emparejadas.....	90
Tabla 22. Pruebas de normalidad Kolmogorov-Smirnov - Shapiro-Wilk	92
Tabla 23. Normalidad de Pvalor	93
Tabla 24. Estadísticas de muestras emparejadas.....	93
Tabla 25. Correlaciones de muestras emparejadas.....	93
Tabla 26. Prueba de muestras emparejadas.....	94
Tabla 27. Definición operacional de las variables	110
Tabla 28. Cuestionario PRE TEST	111
Tabla 29. Cuestionario POST TEST	112
Tabla 30. Opciones de configuración.....	119

RESUMEN

La investigación titulada “Modelo de sistema basado en Web Scraping en la gestión de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja 2019” tiene como objetivo demostrar la influencia que tiene la implementación de un modelo de sistema basado en Web Scraping en la gestión de información en la Unidad de Seguros Públicos y Privados.

En la metodología se utilizó la investigación de tipo aplicada, el nivel de investigación es explicativo y se utilizó el diseño experimental de tipo pre experimental.

Se concluyó que, La implementación de un modelo de sistema basado en Web Scraping influye positivamente en la gestión de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja, al observar que el porcentaje de la media de nivel de información en el Pre Test se obtuvo un 33.60%, en el Post Test se obtuvo 86.80% y la diferencia significativa es de 53.20%; Por lo tanto se afirmó que hay una diferencia significativa en las medias del nivel de gestión de información antes y después de la implementación del modelo de sistema basado en Web Scraping.

Palabras Clave: Web Scraping, gestión de información, Unidad de Seguros Públicos y Privados.

ABSTRACT

The research entitled, “System model based on Web Scraping in information management in the Public and Private Insurance Unit of the Executing Unit 403 Health Network Tayacaja 2019” responds to the research problem: How does the implementation of a system based on Web Scraping influence the quality of information management in the Public and Private Insurance Unit of the Executing Unit 403 Tayacaja Health Network? For this, the objective was to determine how the implementation of a system based on Web Scraping influences the quality of information management in the Public and Private Insurance Unit. It was hypothesized: The implementation of a system based on Web Scraping positively influences the quality of information management in the Public and Private Insurance Unit.

In the methodology applied research was used, the research level is explanatory and the pre-experimental type experimental design was used.

It was concluded that, the implementation of a system based on Web Scraping positively influences the quality of information management in the Public and Private Insurance Unit of the Executing Unit 403 Red de Salud Tayacaja, when observing that the percentage of the mean information quality in the PRE TEST was 33.60%, in the POST TEST it was obtained 86.80% and the significant difference is 53.20%; therefore, it was stated that there is a significant difference in the means of the information management quality level before and after the implementation of the system model based on Web Scraping.

Keywords: Web Scraping, information management, goals, Public and private insurance unit.

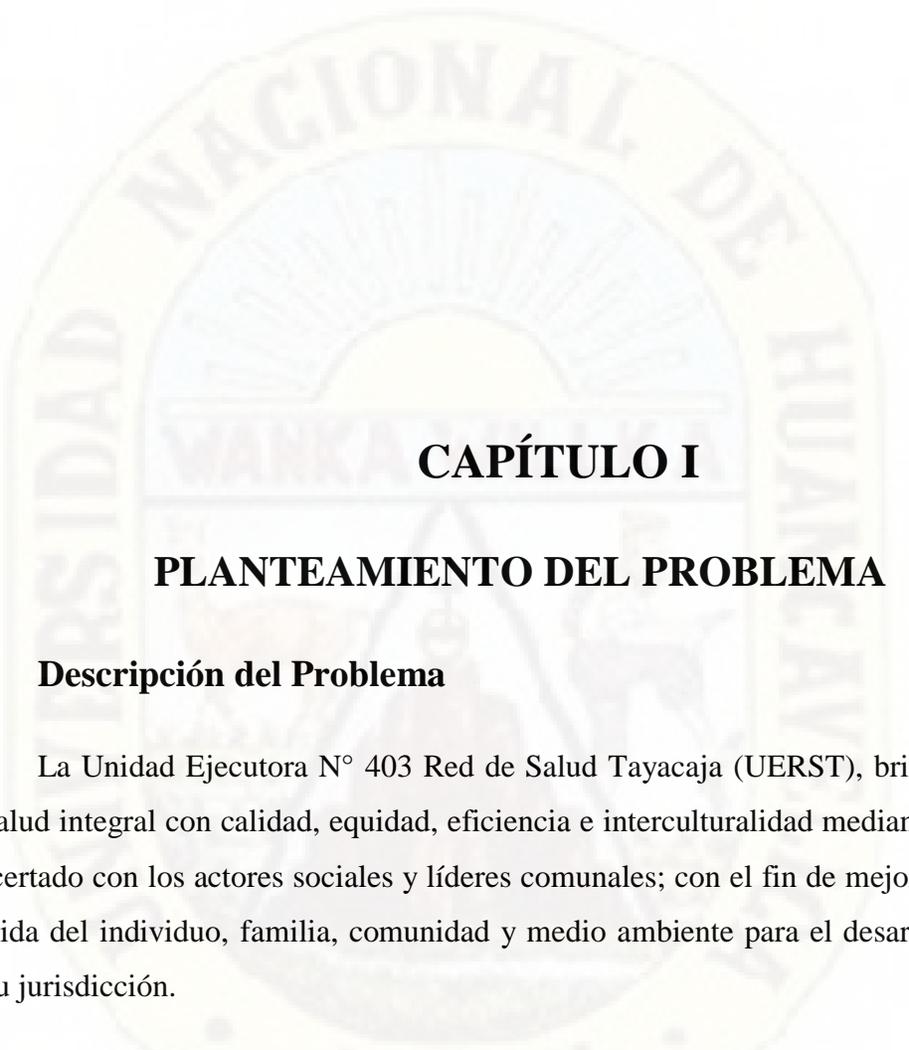
INTRODUCCIÓN

Web Scraping es una técnica que sirve para extraer información estructurada de páginas web de forma automatizada y puede conseguir cantidades industriales de información.

En la presente investigación se desarrolló un modelo de sistema basado en Web Scraping para demostrar la influencia en la gestión de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja, ya que aborda la necesidad de mejorar en la adquisición de información que ayuden a satisfacer a las áreas usuarias y logren tomar decisiones óptimas para el logro de sus metas como institución.

La investigación se ha dividido en cuatro capítulos: El capítulo I; trata sobre el problema en el cual se detalla el planteamiento del problema, formulación del problema, planteamiento del objetivo general, los objetivos y la justificación. El capítulo II; trata sobre el marco teórico donde se expone los antecedentes, las bases teóricas, la hipótesis y las variables de estudio. El capítulo III; se refiere a la metodología de la investigación, en el cual se determina el ámbito de estudio, tipo de investigación, nivel de investigación, método de investigación, diseño de investigación, la población, muestra, técnicas, procedimientos de recolección de datos y el procesamiento de datos. En el capítulo IV, se presenta los resultados. Finalmente, se describen las conclusiones, recomendaciones, referencias bibliográficas y los anexos correspondientes de la presente investigación.

Los tesisistas



CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción del Problema

La Unidad Ejecutora N° 403 Red de Salud Tayacaja (UERST), brinda servicios de salud integral con calidad, equidad, eficiencia e interculturalidad mediante un trabajo concertado con los actores sociales y líderes comunales; con el fin de mejorar la calidad de vida del individuo, familia, comunidad y medio ambiente para el desarrollo integral de su jurisdicción.

Mediante convenio de compra de servicios de salud suscrito entre el Seguro Integral de Salud (SIS) y el Gobierno Regional de Huancavelica (GRH), de fecha 15 de febrero del 2018 modificado por sus adendas suscritas en fecha 30 de mayo de 2018 y 06 de julio de 2018, se establecieron las condiciones bajo las cuales, el GRH y sus Instituciones Prestadoras de Servicios de Salud (IPRESS) dependientes se obligaron a brindar los servicios de salud a los asegurados del SIS, con excepción de aquellos cubiertos por el Fondo Intangible Solidario de Salud (FISSAL) y las atenciones de intercambio prestacional. Así mismo, el SIS se obligó a transferir los fondos al GRH a

través de sus Unidades Ejecutoras por los servicios que las IPRESS presten a los asegurados.

Tabla 1

Población Atendida en un Servicio en Salud por la Micro Red en la USPP de la UERST – Huancavelica, Año 2018

	Red de Salud Tayacaja	MR Acostambo	MR Acraquia	MR Colcabamba	MR Daniel Hernández	MR Pazos	MR San Isidro De A.	MR Surcubamba
AÑO	2018	2018	2018	2018	2018	2018	2018	2018
Afiliados al SIS	79929	10672	8301	13790	18261	10933	6310	11662
Atendidos en el año	58170	8036	5724	10190	12582	7498	4851	9289
%	73	75	69	74	69	69	77	80

Fuente: Oficina de Informática de la Unidad Ejecutora 403 Red de Salud Tayacaja

De los datos obtenidos de la tabla 1, indicamos que los afiliados al aseguramiento universal en salud (AUS) del SIS para el año 2018 son un total de 79 929, de los cuales solo se atendieron o recibieron una atención medica 58 170 pobladores, esto indica que solo el 73 % de la población asegurada recibió una atención en las IPRESS de la UERST.

Tabla 2

Formatos Únicos de Atención (FUA) que no superaron las Reglas de Consistencia del Periodo 2019 Publicados por el SIS

(Actualización de reglas 2019)

PERIODO	DOCUMENTO
2019-12	FUAS que no superaron las reglas de consistencia. Publicado el 2020-01-15 13:00
2019-11	FUAS que no superaron las reglas de consistencia. Publicado el 2020-01-10 13:00
2019-10	FUAS que no superaron las reglas de consistencia. Publicado el 2019-11-22 16:00
2019-09	FUAS que no superaron las reglas de consistencia. Publicado el 2019-11-12 16:00
2019-08	FUAS que no superaron las reglas de consistencia. Publicado el 2019-11-12 16:00
2019-07	FUAS que no superaron las reglas de consistencia. Publicado el 2019-09-30 16:00
2019-06	FUAS que no superaron las reglas de consistencia. Publicado el 2019-09-30 16:00
2019-05	FUAS que no superaron las reglas de consistencia. Publicado el 2019-08-28 17:00
2019-04	FUAS que no superaron las reglas de consistencia. Publicado el 2019-08-20 17:00
2019-03	FUAS que no superaron las reglas de consistencia. Publicado el 2019-08-08 18:42
2019-02	FUAS que no superaron las reglas de consistencia. Publicado el 2019-08-08 18:42
2019-01	FUAS que no superaron las reglas de consistencia. Publicado el 2019-08-08 18:42

Fuente: Oficina de Informática de la Unidad Ejecutora 403 Red de Salud Tayacaja

La tabla anterior nos muestra; por ejemplo, que las observaciones por reglas de consistencia de las atenciones del periodo 2019-01, recién fueron publicados el 08 de

agosto del 2019. Esto indica que después de 6 meses recién se tuvo información para el área usuaria de la USPP.

La siguiente encuesta realizada a los usuarios nos hace entender mejor la problemática que existe en la USPP:

Pregunta 1: ¿Cómo califica la información de las atenciones SIS en la USPP?

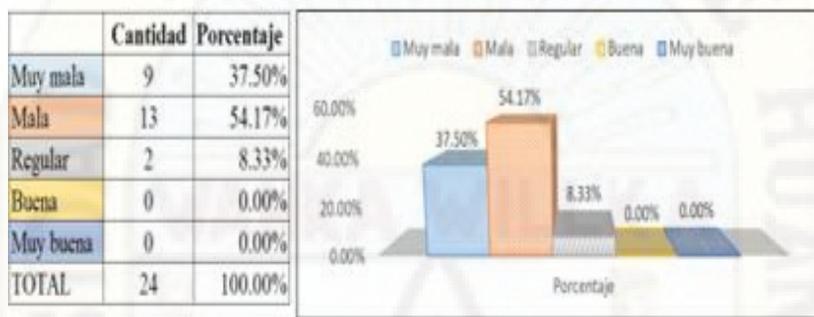


Figura 1. Resultados de la pregunta 01 del PRE TEST

Fuente: Elaboración propia

Estos resultados nos demuestran que los usuarios no están conformes con la información que se distribuye, porque la información de las atenciones SIS es muy deficiente en la USPP.

Pregunta 2: ¿Es eficiente la entrega de la información SIS de la USPP al área usuaria?

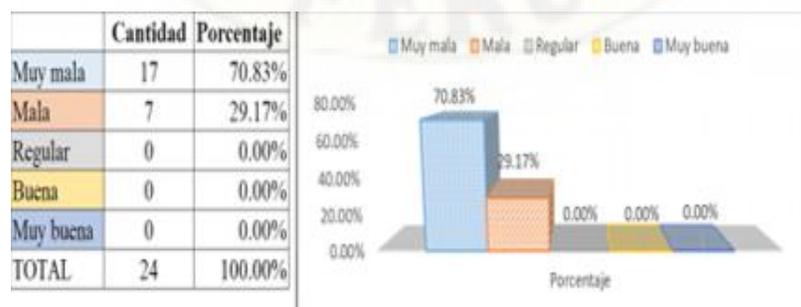


Figura 2. Resultados de la pregunta 02 del PRE TEST

Fuente: Elaboración propia

Según estos resultados los usuarios demuestran que la información que se distribuye no es oportuna.

Pregunta 3: ¿Es comprensible los reportes entregables en la USPP?

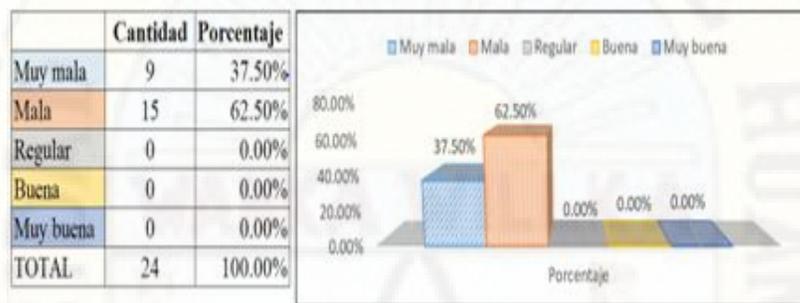


Figura 3. Resultados de la pregunta 03 del PRE TEST

Fuente: Elaboración propia

Estos resultados nos hacen entender que la información que se entrega a los responsables de las áreas no les ayuda a identificar las debilidades o las brechas para el cumplimiento de las metas.

Pregunta 4: ¿El personal de la USPP posee capacidades en el procesamiento de información?

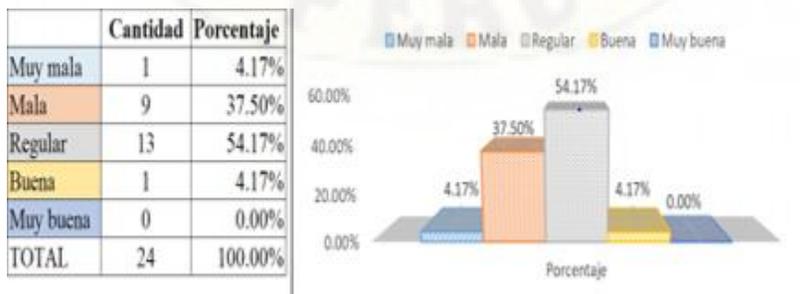


Figura 4. Resultados de la pregunta 04 del PRE TEST

Fuente: Elaboración propia

Esto nos indica que se requiere brindar más asistencia y fortalecer capacidades al personal del área de informática en la adquisición y procesamiento de información.

Pregunta 5: ¿Cómo califica la retroalimentación de la información de las atenciones SIS en la USPP?

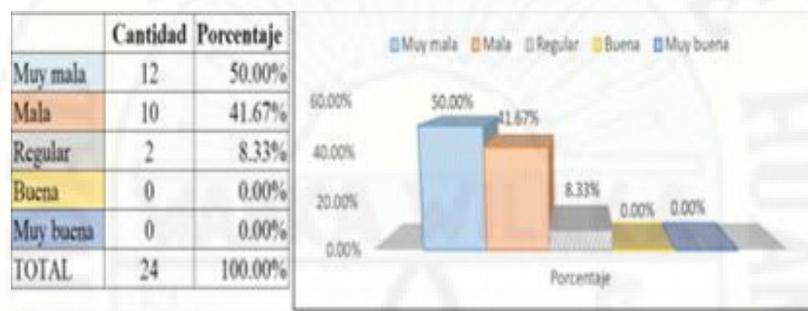


Figura 5. Resultados de la pregunta 05 del PRE TEST

Fuente: Elaboración propia

Esto nos indica que la subsanación del formato único de atención (FUA) observado por reglas de consistencia no es buena; por tanto, se podría determinar que la retroalimentación de la información en la USPP no se está realizando correctamente.

Estos resultados nos indican que en la USPP de la UERST no cuentan con un buen sistema de gestión de información, ya que la información que proporciona a las áreas usuarias es deficiente, la información no es tan buena, la entrega de información no es eficiente, por tal motivo los directivos y responsables de las estrategias no toman decisiones acertadas para el cumplimiento de metas que se maneja en el convenio entre el SIS y la UERST.

1.2. Formulación del Problema

1.2.1. Problema General

¿Cómo influye la implementación de un modelo de sistema basado en Web Scraping en la gestión de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja?

1.2.2. Problemas Específicos

- a) ¿Cómo influye la implementación de un modelo de sistema basado en Web Scraping en la adquisición de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja?
- b) ¿Cómo influye la implementación de un modelo de sistema basado en Web Scraping en la distribución de información de la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja?
- c) ¿Cómo influye la implementación de un modelo de sistema basado en Web Scraping en la toma de decisiones de la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja?

1.3. Objetivos

1.3.1. Objetivo General

Demostrar cómo influye la implementación de un modelo de sistema basado en Web Scraping en la gestión de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

1.3.2. Objetivos Específicos

- a) Demostrar cómo influye la implementación de un modelo de sistema basado en Web Scraping en la adquisición de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

- b) Demostrar cómo influye la implementación de un modelo de sistema basado en Web Scraping en la distribución de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.
- c) Demostrar cómo influye la implementación de un modelo de sistema basado en Web Scraping en la toma de decisiones en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

1.4. Justificación

Seleccionado el tema de la investigación, definido por el planteamiento del problema y establecido los objetivos, se determina las motivaciones que llevan a desarrollar la presente investigación. Estas motivaciones son de carácter teórico, metodológico y práctico.

1.4.1. Justificación Teórica

El estudio en mención utilizó las herramientas y técnicas del Web Scraping para la extracción y obtención de información para luego ser utilizados de acuerdo a los requerimientos del usuario cumpliendo estándares de calidad, los mismos que pueden reemplazar la información general.

Web Scraping utiliza una metodología para el rastreo, análisis, extracción, procesamiento y transformación de información; procesos que cada vez son más segmentados y enfocados al usuario final, generando así la obtención de información consistente, no volátil y confiable para poner a disposición de los usuarios una información pertinente y oportuna. Así mismo se utilizó la metodología ágil scrum, por lo que esta investigación puede ser utilizada como referencia teórica en otro tipo de trabajos similares a nuestro estudio.

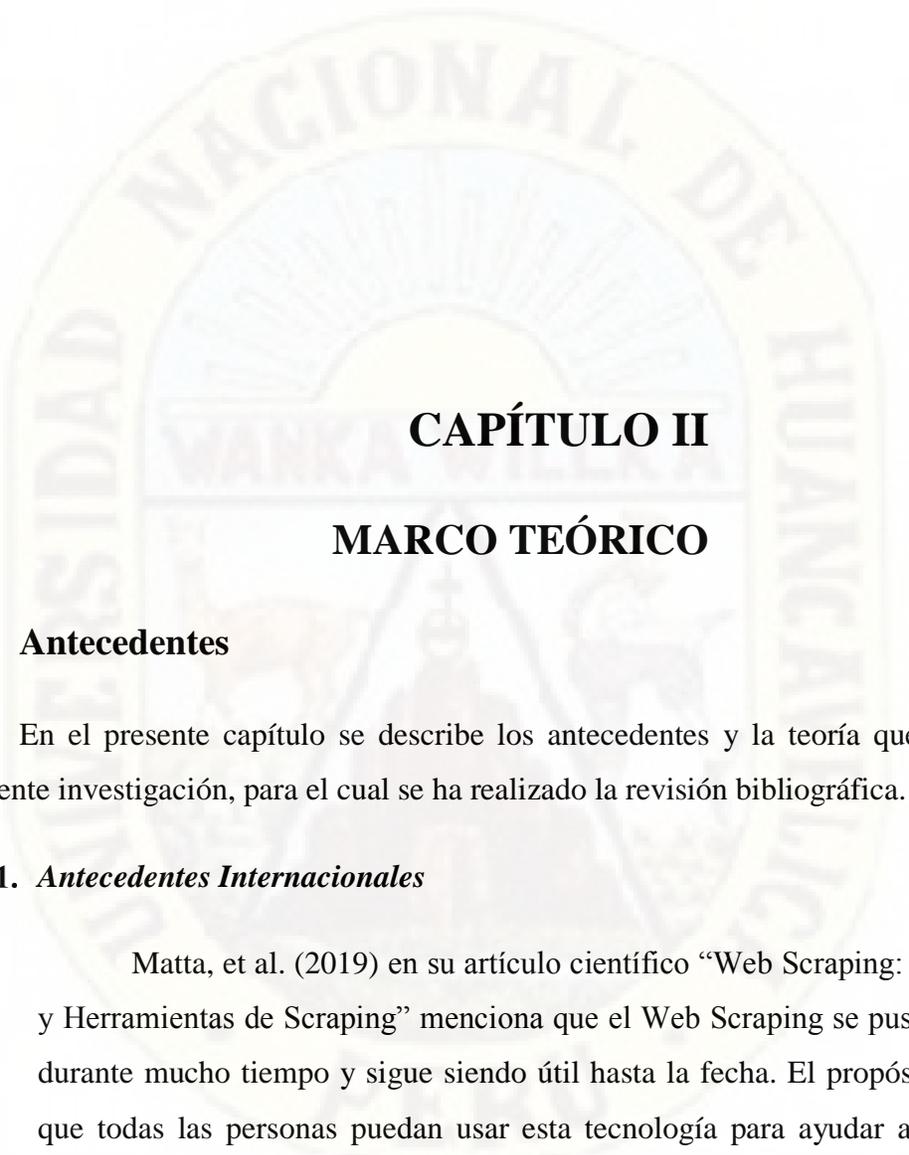
1.4.2. *Justificación Metodológica*

La presente investigación se realiza con fines directamente académicos, con el objetivo de obtener información optima y actualizada el cual servirá para la toma de decisiones de la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

Web Scraping es un método el cual te permite obtener información optima través de los siguientes procesos (rastreo, análisis, extracción, procesamiento y transformación de información), la aplicación de este método será de gran ayuda para demostrar que la forma actual de gestionar información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja – Huancavelica tiene muchas limitaciones y deficiencias, por lo cual no permite obtener la información necesaria para una toma de decisiones correcta. Esto generaría a tener una mejor gestión de información el cual será de mucha importancia para mejorar la gestión y desarrollo de la Red de Salud Tayacaja

1.4.3. *Justificación Práctica*

En toda organización se proyectan a una mejora continua, tal es el caso de la Red de Salud Tayacaja cuyo objetivo es mejorar su gestión de información implementando nuevas estrategias y nuevas formas de gestión que permitan realizar un mejor servicio de salud y así cumplir con todas las metas e indicadores. Con el modelo de Web Scraping se contribuirá significativamente en una mejora continúa proyectada. Ya que con una mejor gestión de información optima se tendrá una mejor toma de decisiones.



CAPÍTULO II

MARCO TEÓRICO

2.1. Antecedentes

En el presente capítulo se describe los antecedentes y la teoría que soportan la presente investigación, para el cual se ha realizado la revisión bibliográfica.

2.1.1. *Antecedentes Internacionales*

Matta, et al. (2019) en su artículo científico “Web Scraping: Aplicaciones y Herramientas de Scraping” menciona que el Web Scraping se puso en práctica durante mucho tiempo y sigue siendo útil hasta la fecha. El propósito fue hacer que todas las personas puedan usar esta tecnología para ayudar a expandir su conocimiento a través de herramientas y aplicaciones relacionadas con el Web Scraping, cuyo resultado es la extracción de gran cantidad de datos para el análisis, clasificación y toma de decisiones. Aporta a la investigación: En el uso de las mejores herramientas a través de la técnica de Web Scraping para la extracción de datos actualizados del web del Ministerio de Salud (MINSA) de las atenciones registradas de los pacientes a través de aplicaciones offline – online

en la Oficina de Unidad de Seguros Públicos y Privados de la Red de Salud Tayacaja.

Diciembre y Luz (2019) en su artículo científico **“Gestión de información en servicios médicos de hospitales”** menciona que el objetivo general es analizar la gestión de información en servicios médicos de hospitales públicos con énfasis en la disposición de tecnologías y sistemas de información, mediante el modelo de evolución de la información planteado por Davis, Miller y Russell (2006) y desde el enfoque de arquitectura tecnológica y la arquitectura de información. Cuya conclusión es la gestión de información en centros de asistencia en salud que comprende la aplicación de tecnologías y su uso en aras de mejorar la asistencia médica y promover una gestión integral de los servicios médicos; el aporte a nuestra investigación radica en que es importante analizar oportunamente la información adecuada y actualizada en aras de mejorar la asistencia médica y consecución de objetivos dentro de la Unidad de Seguros Públicos y Privados de la Red de Salud Tayacaja.

2.1.2. Antecedentes Nacionales

Huaman y Quispe (2019) presentaron su tesis de grado denominado **“Modelo de búsqueda de productos alimenticios en supermercados online categoría abarrotes utilizando asistente virtual de tipo Chatbot y extracción de datos con Web Scraping”**, cuyo objetivo principal fue proponer un modelo de búsqueda de productos alimenticios en supermercados online categoría abarrotes, utilizando asistente virtual de tipo Chatbot y extracción de datos con Web Scraping. El tipo de investigación utilizada es la aplicada, con nivel de investigación correlacional, descriptiva y diseño experimental por qué realizaremos pruebas tomando una muestra de personas que compran en supermercados de Arequipa, las cuáles interactuarán con el modelo para así medir la satisfacción del servicio y el tiempo de búsqueda de productos en

supermercados online. Se llegó a la siguiente conclusión que el Chatbot ofrece un servicio al cliente más personalizado brindándoles información en todo momento, también se determinó que el Chatbot debe estar en constante entrenamiento para brindar respuestas asertivas. En cuanto se pudo determinar que el Web Scraping es una técnica que facilita la extracción de información de forma masiva y automatizada, para que pueda ser usada de acuerdo a las necesidades del usuario.

Este proyecto de investigación guarda relación con este proyecto, por el uso de la metodología Web Scraping y también en la extracción de información automatizada lo cual puede ser utilizada de acuerdo a las necesidades de quien lo requiera.

En el mismo orden Pajuelo Carlevarino (2019) presentó su tesis titulada **“Aplicación web para la gestión de la información de los Programas Sociales en la Municipalidad Provincial del Callao”**, en su trabajo busca resolver este problema mediante la adaptación de una metodología de desarrollo e implementación de una aplicación web para gestionar la información. La metodología seleccionada fue AUP, esta metodología fue adaptada para el trabajo de la gerencia de informática para asegurar que la aplicación web sea de calidad y se adecúe al proceso de la Gerencia General de Programas Sociales. Los resultados obtenidos mediante la operación de variables y realización de encuestas indican que la aplicación web cumple con los indicadores de efectividad, mantenibilidad, usabilidad y disponibilidad propuestos. La implementación de una aplicación web influyó positivamente en la gestión de la información del Programa Comedor del Pueblo de la Gerencia General de Programas Sociales.

El estudio referenciado guarda relación con este proyecto, en que la adaptación de una metodología e implementación de una aplicación web ayuda y

mejora la gestión de información, el mismo que es conforme a lo que se hace en la presente investigación.

Es de resaltar también que la implementación de una aplicación web influyó positivamente en la gestión de la información de la investigación mencionada.

Capuñay (2018) presentó su tesis para optar el grado de magister en la Universidad Nacional “PEDRO RUIZ GALLO” titulada **“ANÁLISIS DEL PERFIL PROFESIONAL DE LOS PROFESIONALES DE INGENIERIA DE SISTEMAS EN INSTITUCIONES DEL SECTOR PUBLICO PERUANO A PARTIR DE LA APLICACIÓN DE TECNICAS DE INTELIGENCIA DE NEGOCIOS Y SCRAPING”**, cuyo objetivo fue identificar las fuentes y realizar la obtención de datos para el análisis del perfil laboral de los profesionales de ingeniería de sistemas para desempeñar cargos en instituciones públicas, a partir de la aplicación de técnicas de scraping de datos. La metodología de investigación utilizada fue de tipo tecnológica – descriptiva. En su investigación logró con éxito identificar la fuente de datos para el análisis del perfil laboral de los profesionales de ingeniería de sistemas para desempeñar cargos en instituciones públicas, la cual corresponde al portal del centro de empleo del Ministerio de Trabajo y Promoción del Empleo de Perú. Asimismo, realizó con éxito la recolección de datos a partir de la aplicación de técnicas de scraping de datos.

El estudio de capuñay se corresponde con la presente investigación específicamente con el uso y aplicación de técnicas de scraping de datos, que es la misma que se utiliza en la presente investigación.

Zuasnabar (2020) presentó su tesis para optar el grado de ingeniero de sistemas y computación en la Universidad Peruana los Andes titulada **“SISTEMA WEB DE CONTROL DE CRECIMIENTO Y DESARROLLO**

DEL NIÑO PARA MEJORAR LA CALIDAD DE ATENCIÓN EN EL HOSPITAL LIRCAY”, cuyo objetivo fue implementar un sistema web de control de crecimiento y desarrollo del niño mediante la metodología XP y medir la mejora en la calidad de atención del Hospital Lircay. El método general de investigación fue el inductivo – deductivo, con enfoque cuantitativo, el tipo de investigación fue aplicada, de nivel Descriptivo – Explicativo y el diseño pre-experimental. La población para este estudio fue de 20 trabajadores de salud, no se tuvo muestreo, debido a que la población fue pequeña. Cuya conclusión fue la implementación del sistema web de control de crecimiento y desarrollo del niño bajo la metodología XP mejora positivamente en la calidad de atención del hospital de Lircay.

Loayza (2015) presento su tesis titulada **“LA GESTIÓN DE LA INFORMACIÓN Y SU RELACIÓN CON LA TOMA DE DECISIONES EN LA DIRECCIÓN DE INFRAESTRUCTURA DE LA UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS ANDAHUAYLAS, 2014”**, cuyo objetivo fue determinar la relación entre la gestión de la información y su relación con la toma de decisiones en la Dirección de Infraestructura de la Universidad Nacional José María Arguedas, mediante un cuestionario estructurado se encuestó a profesionales, técnicos y auxiliares a un total de 44 trabajadores administrativos de la misma universidad. Logrando como resultado demostrar que sí existe relación entre la Gestión de la Información y la Toma de Decisiones en la Dirección de Infraestructura de la Universidad Nacional José María Arguedas Andahuaylas. Y concluye que, la relación determinada es positiva media entre los variables en estudio.

2.2. Bases Teóricas sobre el Tema de Investigación

Para esta investigación y análisis, es de suma importancia definir con exactitud cada uno de los términos que se utilizan en este documento, a fin de evitar malas

interpretaciones de conceptos, es por ello que a continuación se detalla cada uno de los significados más utilizados en este plan.

2.2.1. *Gestión de Información*

La finalidad de la gestión de la información es ofrecer mecanismos que permitieran a la organización adquirir, producir y transmitir, al menor coste posible, datos e informaciones con una calidad, exactitud y actualidad suficientes para servir a los objetivos de la organización. (Flores, 2014).

Salas (2002) menciona que la gestión de información surge como un nuevo concepto dentro del campo de la ciencia de la información, orientado al manejo de la inteligencia corporativa de una organización, que permite la estructuración interna a las organizaciones y les permite reaccionar ante los cambios de su entorno apoyándose en el uso de la información y de los recursos de información disponibles. Esta situación nos lleva a las concepciones más recientes defendidas por la gestión de la información (GI), en el sentido de que las organizaciones deben ser consideradas fundamentalmente como sistemas de información. Algunos autores coinciden en que la gestión de información, plantea que "gerencia es todo lo que se refiere a la obtención de la información adecuada, para la persona adecuada, a su precio adecuado, en el tiempo y lugar adecuado, para tomar la decisión adecuada". Su objetivo es el de incrementar los niveles de eficiencia y efectividad dentro de una organización. Este proceso se conduce a través de la integración adecuada de los recursos humanos, las políticas, las actividades o procedimientos, el hardware, el software y los datos. Owen (1989), denomina la gestión de la información como una triple hélice. En primer lugar, lo importante de la información es su contenido y no tanto su soporte, en segundo lugar, considera que los gastos para sistemas y tecnologías de la información son un gasto para recursos y no deben ser considerados como gastos generales de funcionamiento, y la tercera parte de la filosofía de la gestión de la información es la exacta coordinación del recurso dentro de

la propia organización, ya que en la actualidad este recurso está muy disperso en diversas empresas. (pág. 3)

Los elementos involucrados con la gestión de información se pueden resumir en tres:

- a) Los que competen a la información como fuente/recurso (procesos productivos al interior de las organizaciones)
- b) Los relacionados con el usuario de productos y servicios de información
- c) Los que conforman el canal de comunicación entre el usuario y la fuente.

El objetivo básico de la gestión de la información es garantizar, disponer y organizar la información dentro de la organización los mismos que luego serán distribuidos para la toma de decisiones.

2.2.1.1. Funciones de la Gestión de Información

Arévalo (2007) citada por Rodríguez y Rodríguez (2002) afirma que “las funciones de la gestión de la información abarcarían desde”:

- a) Determinar las necesidades de la información en correspondencia a sus funciones y necesidades.
- b) Mejora de los canales de comunicación y acceso a la información.
- c) Mejora de los procesos informáticos.
- d) Empleo eficiente de los recursos

El mismo autor considera como un recurso, un producto y un activo:

- La información como un recurso tiene un coste y debe tener un rendimiento.
- La información como un producto deber tener unas exigencias de calidad.

- La información como un activo implica que la organización se preocupe en poseerla, gestionarla y utilizarla.

2.2.1.2. *El Ciclo de la Gestión de la Información*

El ciclo de gestión de la información está basado en la valoración de las necesidades de la información de los usuarios, a nivel interno como externo. También se considera como “... un proceso de adaptación de conceptos que incluyen: (i) qué datos deberán recopilarse; (ii) cómo tendrá que procesarse y analizarse los datos; (iii) cómo tendrá que difundirse la información con el fin de apoyar la toma de decisiones, de manera pertinente y con precisión” (Presidencia de la República de Colombia "El Programa Presidencial para la Acción Integral contra Minas Antipersonal (PAICMA)", 2013). Se presenta la estructura general del ciclo de gestión de la información:



Figura 6. El ciclo de la gestión de la información

Fuente: Programa Presidencial para la Acción Integral control Minas Antipersonal – PAICMA- Colombia (2013, P. 16).

2.2.1.3. *Calidad de Información*

La calidad de la información, o conjunto de datos organizados de manera ordenada y útil, puede ser medida por el grado en que se

manifiestan una serie de cualidades (precisa, oportuna, integra y significativa). Otras cualidades de la información son cruciales para que esta pueda ser tratada por medios automáticos de proceso de datos (objetivo y cuantificable). (Núñez, 2013).

2.2.2. Web Scraping

Es un término para varios métodos utilizados para recopilar información de toda la Internet. Por lo general, esto se hace con un software que simula la navegación web humana para recoger trozos especificados de información de diferentes sitios web. En este sentido, cada vez que alguna persona realiza la acción de “copy&paste” por diferentes páginas de la web para obtener datos, lo que ha realizado ha sido un “scrapeo” de datos en la web, y que posteriormente esta información será utilizada para los requerimientos del usuario. Pero esta práctica puede resultar muy costosa al tener que emplear mucho tiempo en el caso de querer obtener muchos datos de diferentes páginas, y posteriormente, organizar y estructurar dichos datos para poder ser utilizados con otra finalidad. (Chris Hanretty, 2013).

Una de las principales ventajas de usar Web Scraping es extraer información de una o varias páginas, para luego poder ordenarla (gestionar esta información) y así poder estructurar en formatos (tablas, gráficos, dashboard,) que sean entendibles y manipulables en pro de mejorar la gestión de información en la organización.

La característica básica de este método es extraer datos web y analizarlos, para luego ser usada por diversas aplicaciones alimentando, automatizando y actualizando base de datos.

2.2.2.1. 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. Esta simulación se realiza mediante programas denominados “bots”, que crean una clonación del clickeo, la lectura y el “copy-paste”, 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.

2.2.2.2. Usos de Web Scraping

Para organizaciones 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 y otros.
- Periodismo de datos.
- Monitoreo a artículos y datos cambiantes.
- Comparación de precios en línea.
- Recolección contactos.
- Investigación.

2.2.2.3. Técnicas para Web Scraping

Las técnicas para Web Scraping varían ampliamente en el esfuerzo y la complejidad, estos métodos son principalmente técnicas especiales utilizadas para encontrar y aislar los elementos de datos dentro del código HTML de una página web. Estas técnicas son realizadas por muchas empresas, aunque algunos consumen mucho tiempo en ser desarrollados y son difíciles de mantener (Michael Keating, 2016). Algunas de estas técnicas son las siguientes:

- **Copiar y pegar:** Esta técnica es la solución más simple para extraer datos de un sitio web, los que hacen este trabajo pueden

seleccionar las secciones de interés de la página web, copiarlos usando atajo de teclado (Ctrl + C) o por medio de un ratón y pegue simultáneamente (Ctrl + V) todo el material necesario en las hojas de datos externos (hojas de cálculo, etc.).

- **Text grepping a través de expresiones regulares:** Esta técnica utiliza el comando unix - grep para extraer información de los sitios web, es una utilidad de línea de comandos que permite realizar búsquedas en texto plano en páginas web que coincidan con una expresión regular.
- **Programación HTTP:** Con esta técnica se puede recuperar información de páginas web tanto estáticas, como dinámicas mediante el envío de peticiones HTTP al servidor web remoto en donde residen los datos. Esto se suele hacer mediante programación de sockets.
- **Análisis de HTML:** Esta técnica permite extraer datos mediante la detección de una secuencia de comandos, plantilla y/o código común en una página web. Esto se lleva a cabo generalmente por uno de los muchos idiomas principales de programación como XQuery, HTQL, Python, PHP, etc.
- **Análisis de DOM (Modelo de Objetos de Documento):** Es la práctica de recuperar el contenido dinámico generado por los scripts del lado del cliente que se ejecutan en un navegador web. Secuencias de comandos del lado del cliente por lo general se incrustan dentro de un documento HTML o XHTML. El contenido dinámico es típicamente formateado en XML que le permite ser transferido desde el sitio web en su formato especificado.
- **Web Scraping software:** Los softwares de Web Scraping normalmente simulan la forma en que los humanos exploran la

web, también transforma el formato inutilizable y no legible de los datos (formato HTML) en utilizable y formato legible (contenido original como texto, imágenes, etc.), que pueden ser fácilmente exportados a hojas de cálculo para su posterior análisis.

2.2.2.4. HTML

Según Cobo (2008) “HTML es un protocolo que establece los requisitos que se deben cumplir para la transferencia de la información, y exige que la información a transferir sea definida mediante un sistema estándar, comprensible para cualquier equipo”. (p. 55)

Por su parte Capuñay (2012) señala que HTML es un lenguaje que permite definir los elementos que conforman una página web”. Además, menciona que “los diseñadores y programadores de páginas web utilizan este lenguaje de modo que los navegadores que utilizamos los usuarios muestren las páginas web después de interpretar su contenido HTML”.

2.2.2.5. Metodología de Implementación

De acuerdo a Alfaro y Gomez (2016) para la implementación de Web Scraping se requiere una metodología que puede dividirse en cuatro etapas que se mencionan a continuación:

- **Extracción de contenido:** En esta etapa accedemos al contenido del sitio web para extraer toda la información contenida en él como texto plano.
- **Análisis del contenido:** Es una de las más importantes ya que debemos identificar la información que queremos extraer. En caso de fallar en la interpretación de los datos, en las siguientes etapas obtendremos información poco relevante para nuestro objetivo.

- **Extracción de información:** En esta etapa accedemos directamente a la información que hemos identificado en la etapa anterior. Extraemos las partes de información que más nos interesan y separamos esta información en partes individuales para su posterior procesamiento.
- **Transformación de información:** Se convierten los datos en un formato manejable para poder almacenarlos, por ejemplo, en una base de datos. Al mismo tiempo, se realiza la modificación de los datos al tipo de dato correcto que facilitará el proceso cuando sea necesario trabajar con los datos. Es decir, si encontramos en el texto un número debemos almacenarlo con el tipo numérico correspondiente (entero, coma flotante, etc.).

2.2.2.6. Herramientas para el Web Scraping

La cantidad de herramientas para obtener información están divididas de acuerdo con el lenguaje de programación que se utilice, tomando en cuenta el nivel de conocimiento en programación tenemos:

a) Usuarios sin conocimientos de programación:

Orientada a este tipo de usuarios, las aplicaciones presentan una interfaz sencilla, mediante el uso de un ratón, pueda realizar todo mediante “Clics”, tratando en lo posible de evitar la programación. Pero al utilizar este tipo de aplicaciones se ven limitados en el rango de obtención de información que un usuario que tiene conocimientos de programación.

- **Web HTTrack:** Permite a un usuario descargar un sitio en un catálogo, donde podrá crear índices, adquirir HTML, imágenes, y documentos. Podemos decir que es un software que permite realizar todas las operaciones de diseño, como

si estuviera en la misma página, también puede actualizar contenido, funciona como una copia externa, donde puede configurarse hasta obtener el resultado.

- **Mozenda:** Utiliza una interfaz sencilla la que puede ser modificada y crear nuevos datos, separar imágenes o datos de cualquier sitio. Ofreciendo también para los especialistas un confort personalizable.
- **WebHarvy:** Una herramienta para realizar scraping web, trabaja con una interfaz sencilla interactuando mediante “Clicks”, para extraer datos, distingue diseños, reconoce frases y para obtener mejores resultados. Puede trabajar con cantidades de datos en conjunto con arreglos SQL (Structures Query Language), CSV, TSV, XML, JSON.
- **Visual web ripper:** Permite ahorrar tiempo trabajando con formatos y un nivel de programación bajo, realizando un trabajo más elaborado o avanzado. También comparte información aplicaciones (CSV, Excel, entre otros.) de fácil organización y observación.

b) Usuarios con conocimientos de programación:

Permite programar al usuario, de manera que puedan ser más personalizadas sus búsquedas de información en la web, el usuario tendrá y analizar la estructura HTML, el código del scraping se hará de acuerdo con la estructura de página a extraer.

- **Beautiful soup:** Es una biblioteca de Python que sirve para analizar la estructura HTML, crea un árbol con los elementos del documento el cual se utiliza para extraer información.

- **Selenium:** Es un entorno de pruebas de software para aplicaciones basadas en la web lo cual se automatiza en el navegador. También se usa para extraer información de la web, analizando primero la estructura HTML y luego llamar a funciones para extraer etiquetas específicas identificadas por algún ID.
- **Scraper.rb:** Aplicación desarrollada en lenguaje de programación Ruby, explora y recopila información que luego es almacenada en una organización XML. Se apoya en las bibliotecas watir y nokogiri. La biblioteca watir controla el comportamiento del programa, pudiendo abrir el sitio web, explorarla mediante “Clics”, y ajustando el documento sin cambiar el código, la biblioteca nokogiri, permite mediante la biblioteca Watir adquirir la información y reunirla en una organización XML.
- **Scraperwiki5:** Para poder componer códigos en lenguaje, Python, Ruby o PHP, no contaba con una interfaz gráfica, por lo que requería tener un conocimiento de programación considerable. Actualmente cuenta con una interfaz sencilla, trabaja con tablas que pueden ser exportadas a Excel o CSV. También incluye consultas de SQL a los datos almacenados. Esta aplicación permite programación de procedimientos en dialectos (C, Clojure, Coffe Script, Lua, Node.js, Perl, PHP, Python, R, Ruby y Shell). También posee alternativas que permiten dividir tablas en archivos PDF, obtener tweets. Y una biblioteca que consolida en Python, pudiendo borrar información localmente.

2.2.2.7. Impedimentos para utilización de Web Scraping

Existen diferentes impedimentos que imposibilitan o dificultan la utilización de Web Scraping, por un lado, aquellos factores que dificultan la realización de Web Scraping, y por otro, las barreras legales que a lo largo de los años se han desarrollado para evitar que se pueda realizar Web Scraping. (Rami Essaid, 2013).

2.2.2.8. Factores que dificultan el uso de Aplicaciones para Web Scraping.

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 20 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.

2.2.2.9. Base de Datos

Las bases de datos son programas que administran información y hacen más ordenada la información, aparte de hacerla fácil de buscar.

Sus características pueden ser ventajosas o desventajosas; pueden ayudarnos para almacenar, organizar, recuperar, comunicar y manejar información en formas que serían imposibles sin los computadores, pero también nos afecta de alguna manera ya que existen enormes cantidades de información en bases de datos de las que no se tiene control del acceso

Las bases de datos tienen muchos usos; nos facilitan el almacenamiento de grandes cantidades de información, permiten la recuperación rápida y flexible de información, con ellas se puede organizar y reorganizar la información, así como imprimirla o distribuirla en formas diversas.

Algunos paquetes comerciales manejadores de bases de datos son: El file Marker Pro, ProQuest Periodical Abstracts, Organizer de Lotus y Far Side Calendar de Amaze.

2.2.2.9.1. SQL

Structured Query Language (SQL), es un lenguaje de programación diseñado específicamente para el acceso a sistemas de gestión de bases de datos relacionales (SGBDR). Este lenguaje es empleado en sistemas informáticos que van desde ordenadores personales muy

básicos con apenas 64 MB de espacio en memoria central hasta los más potentes multiprocesadores y multicomputadoras con decenas de procesadores súper escalares de 64 bits.

Las principales ventajas que aportan SQL son dos:

- Su enorme difusión pues es empleado en la gran mayoría de los sistemas actuales.
- Su elevada expresividad. Por ejemplo, operaciones que costarían semanas de duro esfuerzo en ser desarrolladas en un lenguaje de programación tradicional pueden ser realizadas con SQL en tan sólo unos minutos.

Los usos de SQL incluyen la modificación de las estructuras de tablas e índices de las bases de datos; la adición, actualización y eliminación de filas de datos; y la recuperación de subconjuntos de información de una base de datos para el procesamiento de transacciones y aplicaciones analíticas. Las consultas y otras operaciones SQL adoptan la forma de comandos escritos en forma de sentencias: las sentencias SQL más utilizadas son select, add, insert, update, delete, create, alter y truncate.

2.2.2.9.2. Datos.

El dato es una representación simbólica (numérica, alfabética, algorítmica, espacial, etc.) de un atributo o variable cuantitativa o cualitativa. Los datos describen hechos empíricos, sucesos y entidades. Es un valor o referente que recibe el computador por diferentes medios, los datos representan la información que el programador manipula en la construcción de una solución o en el desarrollo de un algoritmo

2.2.2.9.3. Información.

La información es un conjunto organizado de datos procesados, que constituyen un mensaje que cambia el estado de conocimiento del sujeto o sistema que recibe dicho mensaje.

No existe una única definición de "conocimiento". Sin embargo, existen muchas perspectivas desde las que se puede considerar el conocimiento, siendo la consideración de su función y fundamento, un problema histórico de la reflexión filosófica y de la ciencia. La rama de la filosofía que estudia el conocimiento es la epistemología o teoría del conocimiento. La teoría del conocimiento estudia las posibles formas de relación entre el sujeto y el objeto. Se trata por lo tanto del estudio de la función del entendimiento propia de la persona.

2.2.2.9.4. Registro

Es la información relacionada con una persona, un producto o suceso. En una base de datos un simple archivo es un conjunto de registros.

2.2.2.9.5. Archivo

Es una colección de información relacionada, en él la información se guarda como si fuera un archivero.

2.2.2.9.6. Consulta

Puede ser una búsqueda simple de un registro específico o una solicitud para seleccionar todos los registros que satisfagan un conjunto de criterios.

2.2.2.9.7. Reporte o Informe

Es una lista ordena de los campos y registros seleccionados en un formato fácil de leer.

2.2.3. ISO 25000

La ISO/IEC 25000, también es denominada SQuaRE (System and Software Quality Requirements and Evaluation), un grupo de estándares con finalidad del desarrollo de una base de trabajo de esta manera se dé la evaluación de la calidad del producto software (ISO25000, 2019).

El grupo de la ISO/IEC 25000 es la consecuencia del desarrollo de normas predecesoras, y así específicamente las normas de la ISO/IEC 9126, que desarrollaba las peculiaridades para un modelo de calidad del software, e ISO/IEC 14598, que destacaba la evaluación de productos software. Este grupo de normas ISO/IEC 25000 se encuentra conformada por 5 partes (ISO25000, 2019).

2.3. Bases Conceptuales

2.3.1. Oficina Unidad de Seguros Públicos y Privados

La Unidad de Seguros Públicos y Privados (USPP) es un órgano de línea en la Unidad Ejecutora 403 Red de Salud Tayacaja , encargados de la coordinación para el cumplimiento del "CONVENIO PARA EL FINANCIAMIENTO DE LOS SERVICIOS O PRESTACIONES DE SALUD EN EL PRIMER NIVEL DE ATENCIÓN ENTRE EL SEGURO INTEGRAL DE SALUD Y EL GOBIERNO REGIONAL DE HUANCAVELICA 2018" (Porcentaje de afiliación temprana (DNI en < 30 días), porcentaje de niños/as asegurados al SIS menores de 36 meses suplementado con hierro y dosaje de hemoglobina, niños/as menores de 36 meses afiliados al SIS con diagnóstico de anemia (Dosaje de Hb < 11 mg/dl), porcentaje de niños/as con 13 meses cumplidos y esquema de vacunación completo, porcentaje de desnutrición crónica para niños

menores de 5 años de edad, porcentaje de recién nacidos con 2 controles de crecimiento y desarrollo hasta los 15 días de edad, porcentaje de embarazo en adolescentes, porcentaje de afiliados de 15 años a más tamizados en enfermedades no transmisibles, porcentaje de afiliados al SIS tamizada en problemas de salud mental, extensión de uso de afiliados SIS en atención primaria), entre otros.

2.3.1.1. Funciones de la Oficina de Unidad de Seguros Públicos y Privados

- a) Coordinar para lograr el cumplimiento de los indicadores del "CONVENIO PARA EL FINANCIAMIENTO DE LOS SERVICIOS O PRESTACIONES DE SALUD EN EL PRIMER NIVEL DE ATENCIÓN ENTRE EL SEGURO INTEGRAL DE SALUD Y EL GOBIERNO REGIONAL DE HUANCVELICA 2018"
- b) Supervisar, monitorizar y dar soporte informático a los puntos y centros de digitación de FUAS (formato único de atención)
- c) Recepción, consolidación y procesamiento de información Data del ARFSIS de los PDSIS para la toma de decisiones.
- d) Coordinación, supervisión y monitorizar las prestaciones administrativas (Traslados, sepelio, asignación por casa materna)
- e) Afiliación a la población (personas calificadas como pobres) en aseguramiento universal en salud, para que puedan acceder a un conjunto de prestaciones de salud.
- f) Seguimiento y monitoreo de las rendiciones de caja SIS.
- g) Supervisar y monitorizar a los establecimientos de salud del primer nivel de atención para el cumplimiento de los indicadores del convenio 2017:
 - Porcentaje de afiliación temprana (DNI en < 30 días),
 - Porcentaje de niños/as asegurados al SIS menores de 36 meses suplementado con hierro y dosaje de hemoglobina.

- Niños/as menores de 36 meses afiliados al SIS con diagnóstico de anemia (Dosaje de Hb < 11 mg/dl).
- Porcentaje de niños/as con 13 meses cumplidos y esquema de vacunación completo.
- Porcentaje de desnutrición crónica para niños menores de 5 años de edad.
- Porcentaje de recién nacidos con 2 controles de crecimiento y desarrollo hasta los 15 días de edad,
- Porcentaje de embarazo en adolescentes.
- Porcentaje de afiliados de 15 años a más tamizados en enfermedades no transmisibles.
- Porcentaje de afiliados al SIS tamizada en problemas de salud mental.
- Extensión de uso de afiliados SIS en atención primaria.

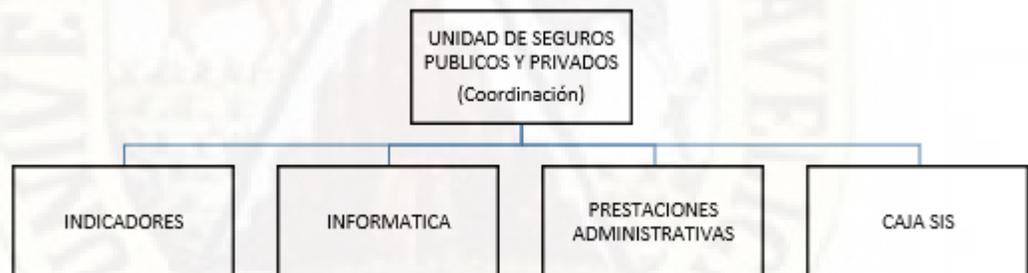


Figura 7. Organigrama de la Oficina de Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja

Fuente: Elaboración propia.

2.3.2. *Web Scraping: ¿Qué programas puedo usar para ser un web scraper?*

Existen varias plataformas y plugins que nos permiten obtener datos de sitios web públicos.

Scraper.io: Es un plugin para Google Chrome muy fácil de usar, aunque su funcionalidad es más limitada que la de WebScraper. Es totalmente gratuito y apenas requiere aprendizaje previo. Permite exportar muy fácilmente los datos a excel.

Import.io: Es una plataforma de integración de datos que proporciona un entorno visual para automatizar el flujo de trabajo de extracción y transformación de datos web. Después de especificar la dirección URL del sitio web de destino, el módulo de extracción de datos web proporciona un entorno visual para diseñar flujos de trabajo automatizados para recopilar datos, yendo más allá del análisis HTML del contenido estático para automatizar las interacciones del usuario final, lo que arroja datos que de otra forma no serían visibles de inmediato.

webhose.io: Es una aplicación web muy potente. Permite realizar 1.000 consultas de manera gratuita y tiene planes de pagos con mayores posibilidades, funcionalidades muy avanzadas y una interfaz muy fácil de usar. (Naranjo, 2017)

Simple HTML DOM Parser: Es útil para manipular los elementos HTML. Esta clase se puede trabajar con ambos valores HTML y HTML que no pasan W3C validación. Usted puede encontrar elementos por identificadores, clases, etiquetas y muchos más. También puede agregar, eliminar o cambiar elementos DOM. La única cosa que usted debe tener en cuenta es la memoria leaks. Pero usted puede evitar memoria leaks. (Makhija, 2015)

En conclusión, sabemos que existen diferentes herramientas o tecnologías que nos permiten extraer a información de diferentes sitios web, es por ello que para recolectar la información de los servicios básicos como son los contactos, horarios de atención, dirección, se hará mediante el uso de esta tecnología.

2.3.3. Información

La información es un conjunto organizado de datos relevantes para uno o más sujetos que extraen de él un conocimiento. Es decir, es una serie de conocimientos comunicados, compartidos o transmitidos y que constituyen por lo tanto algún tipo de mensaje. Sin embargo, su definición varía según la disciplina o el enfoque desde el cual se la piense.

2.3.3.1. Tipos de Información

La información puede clasificarse de maneras muy distintas, conforme a numerosos criterios. Uno de los más comunes tiene que ver con la relación establecida entre los emisores de la información y sus eventuales o posibles receptores, de la siguiente manera:

- **Información Confidencial o Clasificada.** Aquella a la que sólo puede acceder un pequeño conjunto de personas, dada la naturaleza secreta, peligrosa, delicada o privada de los datos contenidos en ella.
- **Información Pública.** Aquella que, por el contrario, permite el acceso general de cualquiera a su contenido, sin requerir permisos especiales y sin tener ningún grado de privacidad.
- **Información Personal.** Aquella que le pertenece a cada persona, es decir, que emana de un individuo concreto, el cual puede decidir con quién compartirla o a quién ofrecérsela.
- **Información Externa.** Aquella que emana de un organismo, institución o empresa, y cuyos destinatarios son instancias o personas externas a la misma.
- **Información Interna.** Aquella, por el contrario, que emana de un organismo, institución o empresa, con el fin de ser consumida de manera interna, sin salir al exterior de la organización.

2.3.3.2. Usos de la Información

La información tiene tantos usos como el receptor pueda darle. Van desde el mero aumento del conocimiento que este posee sobre una materia específica, o quizá del conocimiento aplicable ante una situación determinada, es decir, del conocimiento útil. Leer un manual de uso de un artefacto nos brindará información aplicable de inmediato, por ejemplo.

Otros usos de la información pueden ser más estratégicos, permitiendo al receptor llevar a cabo una mejor toma de decisiones, mejores procesos de control, mejores reglas de evaluación o una mejor elección de alternativas, dependiendo de a qué nos referimos.

Por último, la información tiene un valor cultural, es decir, educativo, divulgativo. A mayor cantidad de información, mayor conocimiento posible, y por lo tanto mayores posibilidades educativas.

2.3.4. Adquisición de Información

Según Chávez (2012) la adquisición de información, son las encargadas de seleccionar y transformar la información desde el ambiente del registro sensorial y de este a la memoria de corto plazo, con ellas se favorece el control y definición de la atención y aquellas que optimizan los procesos de repetición.

2.3.5. Distribución de Información

La distribución de la información implica poner la información necesaria a disposición de los interesados en el proyecto de manera oportuna. La distribución de la información incluye implementar el plan de gestión de las comunicaciones, así como responder a las solicitudes inesperadas de información (Alcántara, 2012).

2.3.6. *Toma de Decisiones*

La toma de decisiones es el proceso mediante el cual se realiza una elección entre alternativas o formas de resolver diferentes situaciones de la vida. ciertos factores pueden influir en las decisiones que tomamos diariamente, a través de algunas técnicas y actividades prácticas. (Universidad de Guadalajara, 2017).

2.4. **Definición de Términos**

- a) **Web Scraping:** Se trata de un proceso de usar bots para extraer contenido y datos de un sitio web, de esta forma se extrae el código HTML y con él, los datos almacenados en la base de datos. Esto supone que se puede duplicar o copiar todo el contenido del sitio web en otro lugar ANTEVENIO, (2019, parr. 1).
- b) **UPPS:** Es la unidad orgánica de apoyo especializado en seguros públicos y privados para efectos de garantizar atenciones de calidad a sus afiliados y para el logro de los objetivos estratégicos y funcionales de la Red de la Red de Salud Tayacaja.
- c) **Gestión de la información:** La gestión de la información se puede identificar como la disciplina que se encargaría de todo lo relacionado con la obtención de la información adecuada, en la forma correcta para la persona indicada, al coste adecuado en el momento oportuno, en el lugar apropiado y articulando todas estas operaciones para el desarrollo de una acción correcta GlossariumBITri, (2009, parr. 1).
- d) **Eficiencia:** Se define como la relación entre los recursos utilizados en un proyecto y los resultados conseguidos con el mismo. Se da cuando se utilizan menos recursos para lograr un mismo objetivo o cuando se logran más objetivos con los mismos o menos recursos. $(\text{recursos planificados}) / (\text{recursos utilizados})$.

Ejemplo: Se es eficiente cuando en 12 horas de trabajo se hacen 100 unidades de un determinado producto. Ahora, se mejora la eficiencia si esas 100 unidades se hacen en solo 10 horas, o se aumenta la eficiencia si en 12 horas se hacen 120 unidades. Aquí se ve que se hace un uso eficiente de un recurso (tiempo), y se logra un objetivo (hacer 100 o 120 productos).

- e) **Eficacia:** Eficacia se refiere al nivel de objetivos conseguidos en un determinado plazo, es decir a la capacidad para conseguir aquello que un grupo se propone. Ser eficaces es simplemente alcanzar la meta estipulada, sin importar el nivel de recursos empleados. $(\text{Resultado alcanzado} * 100) / (\text{Resultado previsto})$

2.5. Hipótesis

2.5.1. Hipótesis General

La implementación de un modelo de sistema basado en Web Scraping influye positivamente en la gestión de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

2.5.2. Hipótesis Específicas

- a) La implementación de un modelo de sistema basado en Web Scraping influye positivamente en la adquisición de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.
- b) La implementación de un modelo de sistema basado en Web Scraping influye positivamente en la distribución de información de la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

- c) La implementación de un modelo de sistema basado en Web Scraping influye positivamente en la toma de decisiones de la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

2.6. Variables

Analizado la hipótesis general y las específicas se ha definido las siguientes variables, y sus dimensiones, la misma que se detalla a continuación:

2.6.1. De la Hipótesis General

La implementación de un modelo de sistema basado en Web Scraping influye positivamente en la gestión de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

2.6.2. Variable Dependiente:

Gestión de información.

2.6.3. Variable Independiente:

Modelo de Sistema basado en Web Scraping

2.7. Operacionalización de Variables

La definición operacional de las variables se describe en el siguiente cuadro:

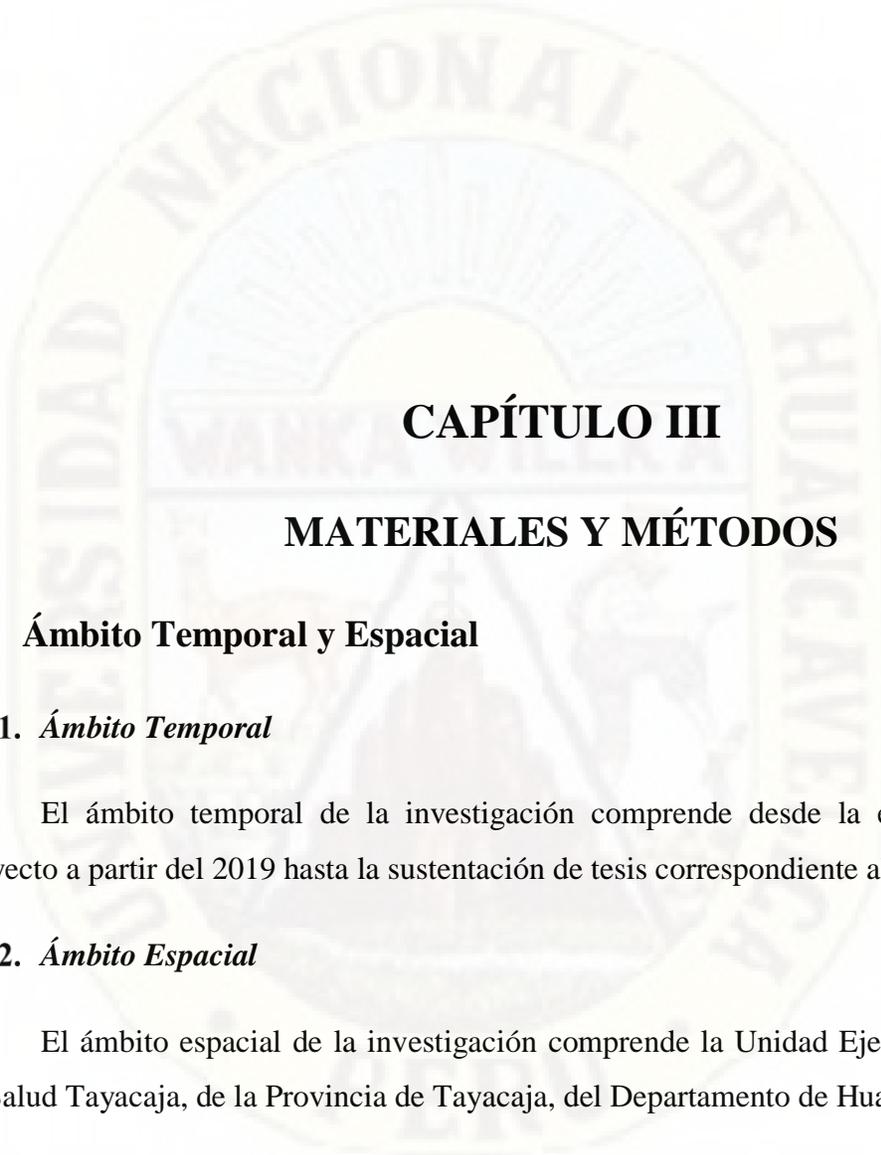
Tabla 3

Definición operacional de las variables

Variable Independiente	Definición	Dimensiones	Indicadores
Modelo de sistema basado en Web	Modelo de sistema basado en Web Scraping, es un método utilizado	Calidad de uso	<ul style="list-style-type: none"> • Funcionalidad • Usabilidad

Scraping	para recopilar información de toda la web, el mismo que permite la satisfacción del usuario y mostrar un sistema adecuado de calidad.	<ul style="list-style-type: none"> • Satisfacción
Variable dependiente	<p data-bbox="665 588 812 630">Definición</p> <p data-bbox="535 840 893 1386">La gestión de la información es considerada como el proceso para la obtención de la información adecuada, en la forma correcta, para la persona u organización indicada, al precio adecuado, en el tiempo oportuno y lugar apropiado, para tomar la decisión adecuada. (Woodman, 1985).</p>	<p data-bbox="974 588 1136 651">Dimensiones</p> <p data-bbox="941 777 1120 892">Adquisición de información</p> <p data-bbox="941 1092 1136 1207">Distribución de información</p> <p data-bbox="958 1396 1120 1470">Toma de decisiones</p> <p data-bbox="1234 598 1404 640">Indicadores</p> <ul style="list-style-type: none"> • Nivel de calidad de información. • Tiempo de adquisición de información. (ítem 1 al 5) • Tiempo de distribución de información. • Nivel de satisfacción de información (ítem 6 al 10) • Tipos • Condiciones • Modelo • Estilos
	Gestión de información	

Elaboración: Propia.



CAPÍTULO III

MATERIALES Y MÉTODOS

3.1. *Ámbito Temporal y Espacial*

3.1.1. *Ámbito Temporal*

El ámbito temporal de la investigación comprende desde la elaboración del proyecto a partir del 2019 hasta la sustentación de tesis correspondiente al año 2021.

3.1.2. *Ámbito Espacial*

El ámbito espacial de la investigación comprende la Unidad Ejecutora 403 Red de Salud Tayacaja, de la Provincia de Tayacaja, del Departamento de Huancavelica.

3.2. *Tipo de Investigación*

La investigación se enmarca dentro del tipo de investigación aplicada.

Es aplicada; o también llamada fáctica, porque su objeto de investigación es una parte de la realidad concreta.

La investigación aplicada concentra su atención en las posibilidades fácticas de llevar a la práctica las teorías generales, y destina sus esfuerzos a resolver los problemas y necesidades de los jefes responsables de todas las estrategias sanitarias en la Unidad Ejecutora 403 Red de Salud Tayacaja, de la provincia de Tayacaja del departamento de Huancavelica,

Según Sampieri (2017) “La investigación aplicada, guarda la relación directa con la básica dependiendo de los descubrimientos, como avances de la investigación básica y a si para enriquecer el conocimiento científico”.

3.3. Nivel de Investigación

Es una investigación de nivel explicativa por que se manipula la variable independiente (sistema basado en Web Scraping) para luego demostrar los efectos producidos en las variables dependiente (gestión de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja).

Según Hernández et al. (2006) “los estudios realizados en la investigación científica, que las investigaciones de nivel explicativa van más allá de la descripción de conceptos, fenómenos. Su interés se centra específicamente en explicar por qué, ocurre un fenómeno natural, en qué, condiciones se manifiestan dichos fenómenos, por qué, se relacionan dos a más variables de diferentes fenómenos naturales”. (pág. 85).

3.3.1. Diseño de Investigación

La investigación realizada pertenece al diseño pre-experimental: con un solo grupo, donde se mide con un Pre-Test y un Post-Test, como se visualiza en la Figura N° 08 diseño de pre prueba y post prueba con un solo grupo.

Según Hernández. R, Fernández. C. Y Baptista. M (2010) el diseño pre-experimental “consiste en administrar un estímulo o tratamiento a un grupo y después

aplicar una medición a una o más variables para observar cuál es el nivel del grupo en éstas”.

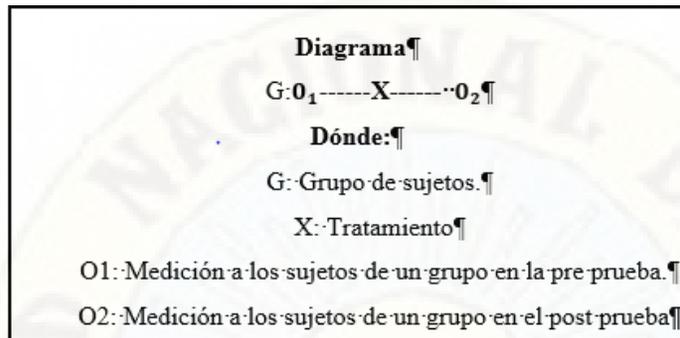


Figura 8. Diseño de pre prueba y post prueba con un solo grupo

3.4. Población, Muestra, Muestreo

3.4.1. Población

La población está constituida por los jefes responsables de las estrategias sanitarias y responsables de las áreas de gestión de la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

La población es un conjunto de individuos de la misma clase, limitada por el estudio. La población se define como la totalidad del fenómeno a estudiar donde las unidades de población poseen una característica común la cual se estudia y da origen a los datos de la investigación.

3.4.2. Muestra

Teniendo como referencia a “La muestra en esencia, es el subgrupo de la población del cual se recolectan los datos y debe ser representativo de esta”

Cuando nuestra población es pequeña, consideramos a todos en el estudio y se le denomina como muestra censal, como dice “sólo cuando queremos realizar un censo

debemos incluir en el estudio a todos los sujetos o casos (personas, animales, plantas, objetos) del universo o la población.

Por lo tanto, la muestra para el proyecto de investigación estará conformada por un total de 24 profesionales entre jefes de área y especialistas en gestión de información de la UERST.

Hallando la muestra censal poblacional

Datos:

N (población)=24

n (muestra)=24

3.5. Técnicas e Instrumentos de Recolección de Datos

Consiste en recolectar datos pertinentes de los atributos, conceptos o variables de las unidades de análisis o haciendo uso de técnicas e instrumentos, como dice “recolectar datos implica elaborar un plan detallado de procedimientos que nos conduzcan a reunir datos con un propósito específico”.

3.5.1. Técnicas

3.5.1.1. Encuesta

Instrumento de modo preferente, en el desarrollo de una investigación: es una técnica ampliamente aplicada de carácter cualitativa.

Se utilizó esta técnica de encuesta con la finalidad de que cada encuestado responda las preguntas en una igualdad de condiciones y sin ninguna presión.

3.5.1.2. Observación

La observación es la técnica de recogida de la información que consiste básicamente, en observar, acumular e interpretar las actuaciones, comportamientos y hechos de las personas u objetos, tal y como las realizan habitualmente.

Se utilizó la ficha de observación, que nos permitió captar comportamientos del sistema basado en Web Scraping.

3.5.2. Instrumentos

Recurso que utiliza el investigador para registrar información o datos sobre las variables.

3.5.2.1. Cuestionarios

Según. “El cuestionario tal vez sea el instrumento más utilizado para recolectar datos, consiste en un conjunto de preguntas respecto de una o más variables a medir.

En esta investigación se utilizó como instrumento el cuestionario que consta de 10 ítems agrupados en 2 dimensiones, de 5 ítems cada uno: adquisición de información y cumplimiento de metas.

3.6. Técnicas y Procesamiento de Análisis de Datos

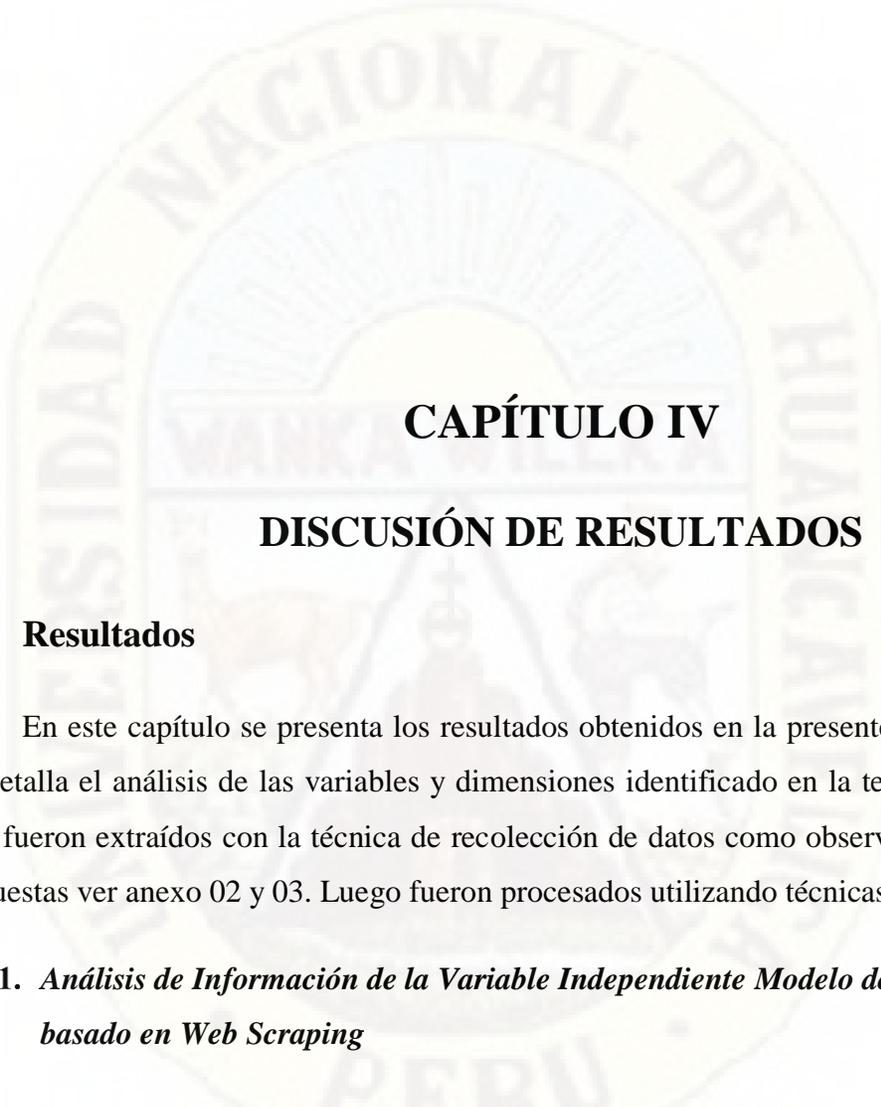
Se realizó el análisis de datos recopilados mediante el cuestionario, aplicados a los responsables de las estrategias sanitarias de la UERST.

Se realizó una evaluación crítica de los datos, a fin de garantizar la verdad y confiabilidad, organizar y procesar en forma secuencial y el estudio de diferentes puntos

de investigación; para a partir de ellos elaborar cuadros estadísticos, gráficos, conceptos técnicos y científicos.

El procesamiento se realizó de manera computarizada empleando el programa SPSS versión 25, posteriormente se plasmó los análisis ligados a la hipótesis.





CAPÍTULO IV

DISCUSIÓN DE RESULTADOS

4.1. Resultados

En este capítulo se presenta los resultados obtenidos en la presente investigación, se detalla el análisis de las variables y dimensiones identificado en la tesis, las mismas que fueron extraídos con la técnica de recolección de datos como observación directa y encuestas ver anexo 02 y 03. Luego fueron procesados utilizando técnicas estadísticas.

4.1.1. Análisis de Información de la Variable Independiente Modelo de Sistema basado en Web Scraping

Para la validación del modelo de sistema basado en Web Scraping se realizó y aplico el modelo de calidad ISO/IEC 25000, conocida como SQuaRE (System and Software Quality Requirements and Evaluation), es una familia de estándares que tiene por objetivo la creación de un marco de trabajo común para así poder evaluar la calidad de sistemas informáticos o software.

La familia del ISO/IEC 25000 tiene como resultado la evolución de las otras normas anteriores, y así especialmente de normas ISO/IEC 9126, que detalla las particularidades de un modelo de calidad del producto software, e ISO/IEC 14598, que se abordaba el proceso de evaluación de productos de software.

Dimensión: Calidad de uso

Indicadores: Funcionalidad, Usabilidad y Satisfacción.

Se utilizó la técnica de recolección de datos como la observación directa y la encuesta. Las mediciones se hicieron utilizando herramientas e instrumentos como: cuestionario, lo mismo que representa las condiciones en las que se encuentran los indicadores en el momento de medición. A continuación, se muestran las pruebas realizadas.

Tabla 4

Resultados obtenidos en la encuesta después de la Implementación del Modelo de Sistema basado en Web Scraping Aplicando la Norma ISO/IEC 25010.

N°	P1	P2	P3	P4	P5	P6
1	1	1	1	1	1	1
2	1	1	1	1	1	1
3	1	1	1	1	1	1
4	0	0	1	1	1	1
5	1	1	1	1	1	1
6	0	0	0	0	0	0
7	1	1	1	1	1	1
8	1	1	1	1	1	1
9	0	0	0	0	0	0
10	1	1	1	1	1	1
11	1	1	1	1	1	1

12	1	1	1	1	1	1
13	0	0	0	0	0	1
14	1	1	1	1	1	1
15	1	1	1	1	1	1
16	0	0	0	0	1	1
17	1	1	1	1	1	1
18	1	1	1	1	1	1
19	1	1	1	1	1	1
20	0	0	1	0	1	1
21	1	0	1	1	1	1
22	1	1	1	1	1	1
23	1	1	1	1	1	1
24	0	0	0	0	1	1
TOTAL	17	16	19	18	21	22

Tabla 5

Resumen del Modelo de la Lista de Cotejo para el Modelo de Sistema basado en Web Scraping.

Calidad de uso			
Funcionalidad		SI	NO
1. Proporciona información solicitada		17	07
2. Cumple las tareas y objetivos del usuario		16	08
	Promedio	16,5	7,5
Usabilidad			
3. Es útil		19	5
4. Es sencillo de utilizar		18	6

	Promedio	18,5	5,5
Satisfacción			
5. Permite realizar menos procesos para obtener la información		21	3
6. Optimiza el tiempo en la obtención de información		22	2
	Promedio	21,5	2,5

Análisis e Interpretación

Al analizar los resultados en la **Tabla N° 5**, se obtiene los siguientes resultados correspondientes a la calidad de uso del modelo de sistema basado en Web Scraping mediante el cuestionario; en el indicador funcionalidad se obtiene un resultado de 16,5 sobre un total de 24 usuarios, del indicador usabilidad fue 18,5 de un total de 24 usuarios y por último del indicador satisfacción el resultado fue 21,5 de un total de 24 usuarios. Durante todas las pruebas realizadas por los usuarios, el modelo del sistema basado en Web Scraping demostró efectividad permitiendo obtener la información oportuna para la toma de decisiones y consecución de metas según los indicadores de gestión del área usuaria. Esto demuestra, la calidad de uso del modelo de sistema según la norma ISO 9126.



Figura 9. Grafica de la calidad de uso del modelo de sistema basado en Web Scraping

Fuente: Elaboración propia

4.1.2. Análisis de Información de la Variable Dependiente Gestión de Información

Para realizar el análisis de la variable dependiente, se procedió a evaluar los indicadores de las dimensiones siguiendo la operacionalización de la variable.

4.1.1. Dimensión: Adquisición de Información

Los indicadores: Tiempo de adquisición de datos (días) y tiempo de procesamiento de datos (días).

Tabla 6

Resultados de la encuesta realizada antes y después de la Implementación del Modelo de Sistema basado en Web Scraping en la Adquisición de Información.

Responsables (área usuaria)	Adquisición de información PRE TEST		Adquisición de información POST TEST	
	Tiempo de adquisición de datos (días)	Tiempo de procesamiento de datos (días)	Tiempo de adquisición de datos (días)	Tiempo de procesamiento de datos (días)
1	20	15	2	6
2	22	20	1	4
3	18	21	1	1
4	24	22	2	3
5	21	16	3	6
6	23	20	2	5
7	26	18	2	5
8	24	19	4	4

9	19	17	3	2
10	25	22	2	2
11	25	23	1	3
12	28	19	2	3
13	20	26	3	3
14	27	18	3	1
15	24	17	1	4
16	26	21	1	5
17	28	18	2	5
18	30	23	1	4
19	18	19	2	6
20	25	16	2	2
21	24	14	3	3
22	24	15	4	3
23	28	13	2	2
24	20	22	1	2

Tabla 7

Resultados de frecuencias obtenidas antes y después de la Implementación del Modelo de Sistema basado en Web Scraping en la Adquisición de Información.

Preguntas	escala	GRUPO EXPERIMENTAL			
		PRE PRUEBA		POST PRUEBA	
		fi	%	Fi	%
Tiempo de adquisición de datos (días)	0 a 7 días	0	0%	24	100%
	8 a 14 días	0	0%	0	0%

	15 días a más	24	100%	0	0%
	Σ	24	100%	24	100%
	0 a 7 días	0	0	24	100%
	8 a 14 días	2	8%	0	0%
Tiempo de procesamiento de datos (días)	15 días a más	22	92%	0	0%
	Σ	24	100%	24	100%

Análisis e Interpretación:

Al analizar los resultados en la **Tabla N° 7**, en una prueba realizada a 24 responsables de las áreas se observa que; antes de la implementación del modelo de sistema Web Scraping, el tiempo de adquisición de datos era de 15 días a más y el tiempo de procesamiento de datos era de 15 días a más; Y con la implementación del sistema Web Scraping el tiempo de adquisición de datos se redujo en el tiempo de 0 a 7 días y el tiempo de procesamiento se redujeron de entre 0 a 7 días; Por lo tanto, se demuestra que con la implementación de un modelo de sistema basado en Web Scraping influye positivamente en la adquisición de Información en la USPP de la UERST.

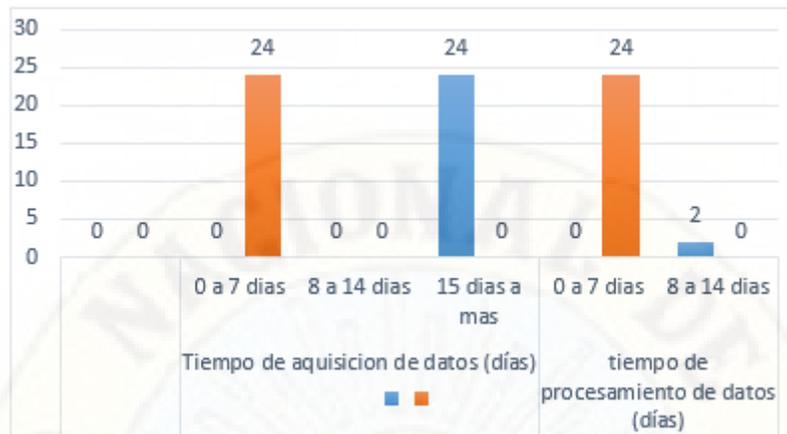


Figura 10. Resultado total de respuestas y promedio del tiempo de adquisición de información

Fuente: Elaboración Propia

4.1.2. Dimensión: Distribución de Información

Los indicadores: Tiempo de distribución de información (días) y tiempo de impresión de reportes (días)

Tabla 8

Resultados de la encuesta realizada antes y después de la Implementación del Modelo de Sistema basado en Web Scraping en la Distribución de Información.

Responsables (área usuaria)	Distribución de información PRE TEST		Distribución de información POST TEST	
	Tiempo de distribución de información (días)	Tiempo de impresión de reportes (días)	Tiempo de distribución de información (días)	Tiempo de impresión de reportes (días)
1	8	12	1	2
2	5	14	3	5
3	7	12	2	2
4	6	15	1	3

5	10	8	2	3
6	5	12	3	1
7	8	13	2	3
8	9	14	2	4
9	10	12	1	1
10	7	8	2	1
11	8	11	1	2
12	7	14	2	1
13	6	8	1	5
14	9	15	2	3
15	11	11	1	2
16	10	10	3	1
17	9	9	2	3
18	8	14	1	1
19	9	15	2	3
20	6	14	2	2
21	7	9	1	1
22	5	12	2	3
23	6	14	2	4
24	7	8	1	2

Tabla 9

Resultados de frecuencias obtenidas antes y después de la Implementación del Modelo de Sistema basado en Web Scraping en la Distribución de Información.

preguntas	escala	GRUPO EXPERIMENTAL	
		PRE PRUEBA	POST PRUEBA

		Fi	%	fi	%
Tiempo de distribución de información (días)	0 a 7 días	12	50%	24	100%
	8 a 14 días	12	50%	0	0%
	15 días a mas	0	0%	0	0%
	Σ	24	100%	24	100%
Tiempo de impresión de reportes (días)	0 a 7 días	0	0%	24	100%
	8 a 14 días	21	87.5%	0	0%
	15 días a mas	3	12.5%	0	0%
	Σ	24	100%	24	100%

Análisis e Interpretación:

Al analizar los resultados en la **Tabla N° 9**, en una prueba realizada a 24 responsables de las áreas se observa que; antes de la implementación del modelo de sistema Web Scraping, el tiempo de distribución de información era en un 50% de 0 a 7 días y en un 50% de 8 a 14 días y el tiempo de impresión de reportes la demora era en un 87.5 % de 8 a 14 días; y con la implementación del sistema Web Scraping se redujo en el tiempo de 0 a 7 días la distribución de información, Por tanto se demuestra que con la implementación de un modelo de sistema basado en Web Scraping influye positivamente en la distribución de información en la USPP de la UERST.

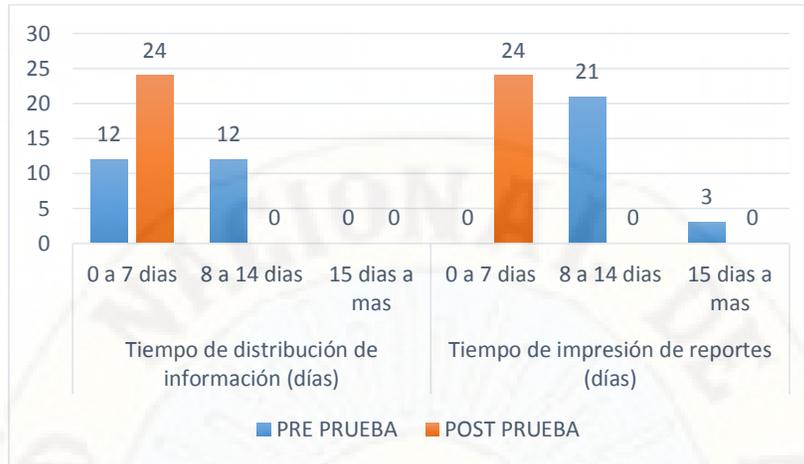


Figura 11. Resultado total de respuestas y promedio del tiempo de distribución de información

Fuente: Elaboración Propia

4.1.3. Dimensión: Toma de Decisiones

Los indicadores: Tipo de información, Condiciones para la toma de decisiones, modelo de toma de decisiones y estilos de toma de decisiones.

Para realizar la evaluación en este indicador, se aplicó una encuesta al área usuaria que son 24 profesionales responsables de las estrategias sanitarias, con 5 rangos de escala de Likert, donde:

EN DESACUERDO (ED)	POCO DE ACUERDO (PA)	DE ACUERDO (DA)	MUY DE ACUERDO (MA)	TOTALMENDE DE ACUERDO (TA)
1	2	3	4	5

Figura 12. Escala de Likert

Tabla 10

Resultados de la encuesta realizada antes y después de la implementación del modelo de sistema basado en Web Scraping en la toma de decisiones.

Toma de Decisiones PRE TEST				Toma de Decisiones POST TEST				
¿La toma de decisiones, está en función del tipo de la información que se utiliza en la oficina?	¿Dispone				¿Dispone			
	s con				s con			
	información completa y confiable, seguro de lo que ocurrirá sobre la alternativa	¿Realizas actividades con información completa y perfecta, eliminando la incertidumbre que maximice el logro de las metas de la oficina?	¿Buscas información de calidad, cantidad y momento oportuno para solucionar problemas complejos?		información completa y confiable, seguro de lo que ocurrirá sobre la alternativa	¿Realizas actividades con información completa y perfecta, eliminando la incertidumbre que maximice el logro de las metas de la oficina?	¿Buscas información de calidad, cantidad y momento oportuno para solucionar problemas complejos?	
	seleccionada, con pocas posibilidades de equivocarse?				seleccionada, con pocas posibilidades de equivocarse?			
	Condiciones para la toma de decisiones	Modelo de toma de decisiones	Estilos de toma de decisiones		Condiciones para la toma de decisiones	Modelo de toma de decisiones	Estilos de toma de decisiones	
				Tipo de información				Tipo de información
	2	2	1	2	3	5	4	4
	1	2	2	1	4	3	4	4
	1	2	3	2	4	3	3	5

2	1	1	1	3	3	4	4
3	1	2	2	3	5	3	3
1	2	2	1	4	5	4	4
1	2	2	3	3	4	5	4
2	3	1	2	5	4	5	3
1	2	3	2	3	4	4	4
2	1	2	2	4	5	4	3
4	2	3	1	3	4	3	4
2	1	1	1	5	3	5	5
2	2	2	1	5	3	4	5
1	2	2	2	3	4	4	4
2	2	1	2	4	4	5	5
2	1	2	1	5	5	4	4
1	1	1	2	3	4	4	5
2	2	1	3	5	5	3	4
1	1	2	3	4	4	4	3
2	1	2	1	5	4	3	4
2	2	1	2	3	5	4	5
3	3	2	1	4	4	3	4
2	1	1	2	5	4	5	3
2	2	1	2	4	3	5	4

Tabla 11

Resultados de frecuencias obtenidas antes y después de la implementación del modelo de sistema basado en Web Scraping en la toma de decisiones.

		GRUPO EXPERIMENTAL			
preguntas	escala	PRE PRUEBA		POST PRUEBA	
		fi	%	fi	%
¿La toma de decisiones, está en función del tipo de la información que se utiliza en la oficina?	EN DESACUERDO (ED)	8	33%	0	0%
	POCO DE ACUERDO (PA)	13	54%	0	0%
	DE ACUERDO(DA)	2	8%	9	38%
	MUY DE ACUERDO (MA)	1	4%	8	33%
	TOTALMENDE DE ACUERDO (TA)	0	0%	7	29%
	Σ	24	100%	24	100%
¿Dispones con información completa y confiable, seguro de lo que ocurrirá sobre la alternativa seleccionada, con pocas posibilidades de equivocarse?	EN DESACUERDO (ED)	9	38%	0	0%
	POCO DE ACUERDO (PA)	13	54%	0	0%
	DE ACUERDO(DA)	2	8%	6	25%
	MUY DE ACUERDO (MA)	0	0%	11	46%
	TOTALMENDE DE ACUERDO (TA)	0	0%	7	29%

	Σ	24	100%	24	100%
¿Realizas actividades con información completa y perfecta, eliminando la incertidumbre que maximice el logro de las metas de la oficina?	EN DESACUERDO (ED)	10	42%	0	0%
	POCO DE ACUERDO (PA)	11	46%	0	0%
	DE ACUERDO(DA)	3	13%	6	25%
	MUY DE ACUERDO (MA)	0	0%	12	50%
	TOTALMENDE DE ACUERDO (TA)	0	0%	6	25%
	Σ	24	100%	24	100%
¿Buscas información de calidad, cantidad y momento oportuno para solucionar problemas complejos?	EN DESACUERDO (ED)	9	38%	0	0%
	POCO DE ACUERDO (PA)	12	50%	0	0%
	DE ACUERDO(DA)	3	13%	5	21%
	MUY DE ACUERDO (MA)	0	0%	13	54%
	TOTALMENDE DE ACUERDO (TA)	0	0%	6	25%
	Σ	24	100%	24	100%

Análisis e Interpretación: Tipo de información

Al analizar los resultados en la **Tabla N° 11**, en una encuesta realizada a 24 responsables de las áreas se observa que; antes de la implementación del modelo de sistema Web Scraping, un 54% y 33% estaba poco de acuerdo y en desacuerdo al tipo de información para la toma de decisiones respectivamente; y con la implementación del sistema Web Scraping un 38% y un 33% estuvo de acuerdo y muy de acuerdo con el tipo de información que se gestionaba para la toma de decisiones. Por lo tanto, se demuestra que con la implementación de un modelo de sistema basado en Web Scraping influye positivamente en la toma de decisiones en la USPP de la UERST.

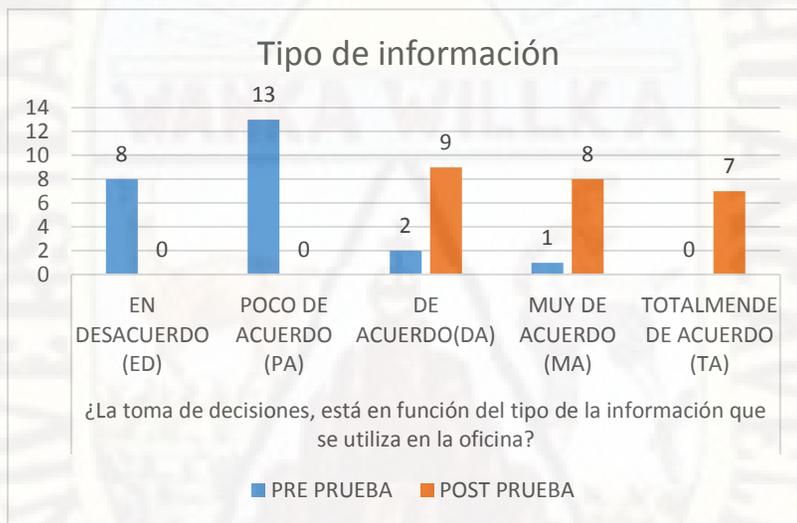


Figura 13. Resultado de respuestas y promedio de la toma de decisiones

Fuente: Elaboración Propia

Análisis e Interpretación: Condiciones para la toma de decisiones

Al analizar los resultados en la **Tabla N° 11**, en una encuesta realizada a 24 responsables de las áreas se observa que; antes de la implementación del modelo de sistema Web Scraping, solo un 8% estaba de acuerdo con la disposición de información completa y confiable para la toma de decisiones. Y con la implementación del sistema Web Scraping, un 25% y un 46% estuvo de acuerdo y muy de acuerdo con la disposición de información completa y confiable para la toma de decisiones. Por lo

tanto, se demuestra que con la implementación de un modelo de sistema basado en Web Scraping influye positivamente en la toma de decisiones en la USPP de la UERST.

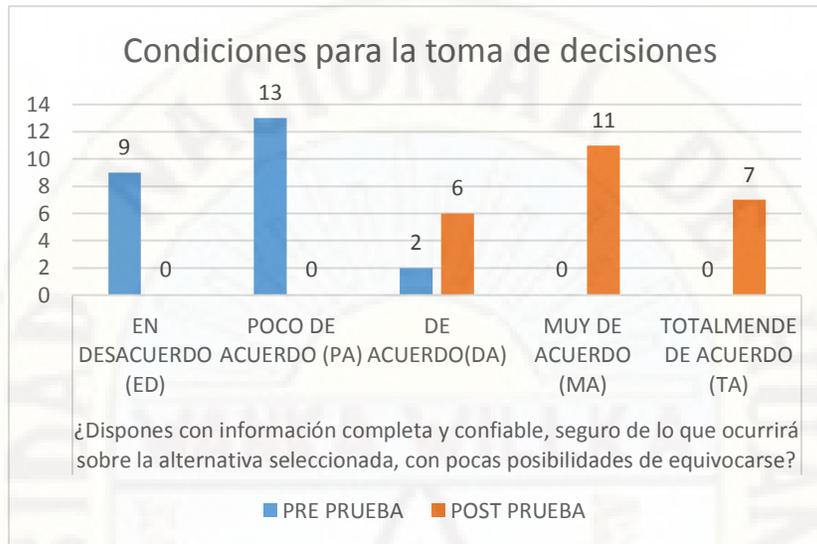


Figura 14. Resultado total de respuestas y promedio de condiciones para la toma de decisiones

Fuente: Elaboración Propia

Análisis e Interpretación: Modelo de toma de decisiones

Al analizar los resultados en la **Tabla N° 11**, en una encuesta realizada a 24 responsables de las áreas se observa que; antes de la implementación del modelo de sistema Web Scraping, solo un 13% estaba de acuerdo con la realización de actividades con la información completa y perfecta para el logro de las metas de la oficina; y con la implementación del sistema Web Scraping, un 50% estuvo muy de acuerdo con la realización de actividades con la información completa y perfecta para el logro de las metas de la oficina. Por lo tanto, se demuestra que con la implementación de un modelo de sistema basado en Web Scraping influye positivamente en la toma de decisiones en la USPP de la UERST.

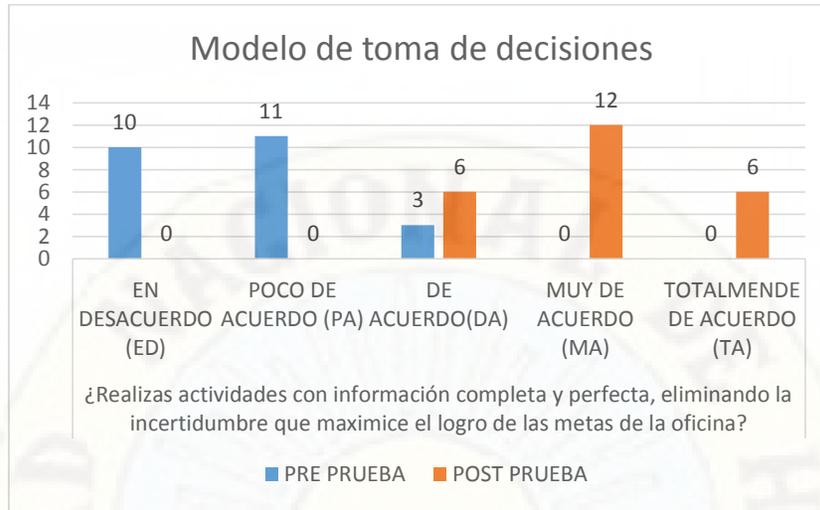


Figura 15. Resultado total de respuestas y promedio del modelo de toma de decisiones

Fuente: Elaboración Propia

Análisis e Interpretación: Estilos de toma de decisiones

Al analizar los resultados en la **Tabla N° 11**, en una encuesta realizada a 24 responsables de las áreas se observa que; antes de la implementación del modelo de sistema Web Scraping, un 50% estaba poco de acuerdo con la búsqueda de información de calidad, cantidad y momento oportuno para solucionar problemas complejos; y con la implementación del sistema Web Scraping, un 54% estuvo muy de acuerdo con la búsqueda de información de calidad, cantidad y momento oportuno para solucionar problemas complejos. Por lo tanto, se demuestra que con la implementación de un modelo de sistema basado en Web Scraping influye positivamente en la toma de decisiones en la USPP de la UERST.

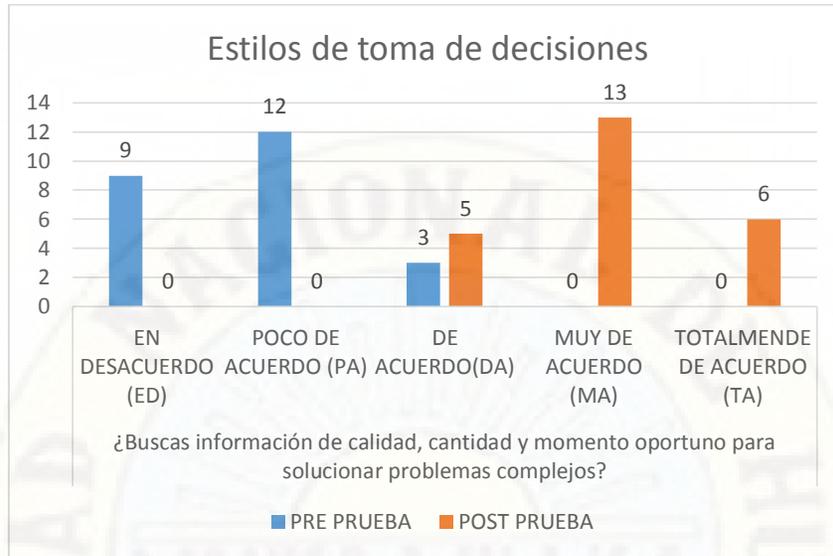


Figura 16. Resultado total de respuesta y promedio de estilos de toma de decisiones

Fuente: Elaboración Propia

4.2. Prueba de Hipótesis

4.2.1. Prueba de la Hipótesis Específica N° 01

La implementación de un sistema basado en Web Scraping influye positivamente en la adquisición de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

4.2.1.1. Planteamiento de las Hipótesis

Ho: El uso de un sistema basado en Web Scraping NO influye positivamente en la adquisición de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja

Ha: El uso de un sistema basado en Web Scraping influye positivamente en la adquisición de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja

4.2.1.2. *Determinación del Nivel de Significancia:*

- Alfa = 95%.
- Error= 0.05

4.2.1.3. *Elección de la Prueba Estadística*

Se elige la prueba de T Student, para muestras relacionadas, por ser un estudio longitudinal con medidas del antes y después.

4.2.1.4. *Cálculo del Valor Tabular:*

NORMALIDAD:

Se elige la prueba de Shapiro Wilk (<33), porque nuestra muestra es menor a 33.

Criterios para determinar la Normalidad:

P valor \Rightarrow α Aceptar H_0 = Los datos provienen de una distribución normal

P valor $<$ α Aceptar H_1 = Los datos NO provienen de una distribución normal

Tabla 12

Pruebas de normalidad Kolmogorov-Smirnov - Shapiro-Wilk

Pruebas de normalidad						
Kolmogorov-Smirnov ^a			Shapiro-Wilk			
Estadístico	Gl	Sig.	Estadístico	gl	Sig.	
o			o			

Promedio_Tiempo_Adquisició n_informacion_PRE	0.109	24	,200*	0.970	24	0.665
Promedio_Tiempo_Adquisició n_información_POST	0.171	24	0.068	0.960	24	0.445

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Se observa el P valor:

Tabla 13
Normalidad de Pvalor

NORMALIDAD			
P valor (PRE PRUEBA)= 0.665	>	$\alpha = 0.05$	
P-valor (POST PRUEBA) = 0.445	>	$\alpha = 0.05$	

Conclusión: Los datos de la dimensión adquisición de Información provienen de una distribución normal.

Decisión Estadística

Tabla 14
Estadísticas de muestras emparejadas

Estadísticas de muestras emparejadas			
Media	N	Desv.	Desv.

				Desviación	Error
				promedio	
Pa r 1	Promedio_Tiempo_Adquisición_informacion_ PRE	21.312	2	2.23029	0.45526
	Promedio_Tiempo_Adquisición_información_ POST	2.7917	2	0.88363	0.18037

Tabla 15

Correlaciones de muestras emparejadas

Correlaciones de muestras emparejadas

		N	Correlación	Sig.
Par 1	Promedio_Tiempo_Adquisición_informacion_ PRE & Promedio_Tiempo_Adquisición_información_ POST	24	-0.285	0.176

Tabla 16

Prueba de muestras emparejadas

Prueba de muestras emparejadas

		t	gl	Sig.
Diferencias emparejadas				(bilateral)

4.2.2. Prueba de la Hipótesis Especifica N° 02

La implementación de un sistema basado en Web Scraping influye positivamente en la distribución de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

4.2.2.1. Planteamiento de las Hipótesis:

Ho: El uso de un sistema basado en Web Scraping NO influye positivamente en la distribución de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

Ha: El uso de un sistema basado en Web Scraping influye positivamente en la distribución de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

4.2.2.2. Determinación del Nivel de Significancia:

- Alfa = 95%.
- Error= 0.05

4.2.2.3. Elección de la Prueba Estadística

Se elige la prueba de T Student, para muestras relacionada, por ser un estudio longitudinal con medidas del antes y después.

4.2.2.4. Cálculo del Valor Tabular:

NORMALIDAD:

Se elige la prueba de Shapiro Wilk (<33), porque nuestra muestra es menor a 33.

4.2.2.5. Criterios para determinar la Normalidad:

P valor $\Rightarrow \alpha$ Aceptar H_0 = Los datos provienen de una distribución normal

P valor $< \alpha$ Aceptar H_1 = Los datos NO provienen de una distribución normal

Tabla 17

Pruebas de normalidad Kolmogorov-Smirnov - Shapiro-Wilk

Pruebas de normalidad							
	Kolmogorov-Smirnov ^a			Shapiro-Wilk			
	Estadístico	gl	Sig.	Estadístico	gl	Sig.	
Promedio_tiempo_distribución_información_PRE	0.118	24	,200*	0.967	24	0.583	
Promedio_tiempo_distribución_información_POST	0.158	24	0.127	0.933	24	0.111	

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Se observa el P valor:

Tabla 18

Normalidad de Pvalor

Normalidad

P valor (Pre Prueba) = 0.583 > $\alpha = 0.05$

P-valor (Post Prueba) = 0.111 > $\alpha = 0.05$

Conclusión: Los datos de la dimensión cumplimiento de metas provienen de una distribución normal.

Tabla 19

Estadísticas de muestras emparejadas

Estadísticas de muestras emparejadas

	Medi a	N	Desv. Desviació n	Desv. Error promedi o
Pa r 1 PRE	Promedio_tiempo_distribución_información_ 9.729 2	24	1.39860	0.28549
POST	Promedio_tiempo_distribución_información_ 2.083 3	24	0.74697	0.15248

Tabla 20

Correlaciones de muestras emparejadas

Correlaciones de muestras emparejadas

N	Correlación	Sig.
---	-------------	------

Par	Promedio_tiempo_distribución_información_PRE			
&		24	0.023	0.917
1	Promedio_tiempo_distribución_información_POST			

Tabla 21
Prueba de muestras emparejadas

		Prueba de muestras emparejadas							
		Diferencias emparejadas							
		Med	Desv.	Desv.	Desv.	95% de			
		ia	Desvi	Error	prom	intervalo	t	g	Sig.
			ación	prom	edio	de		1	(bilat
				edio		confianza			eral)
						de la			
						diferencia			
							Infe	Sup	
							rior	erior	
P	Promedio_tiempo_distribuci								
a	ón_información_PRE -	7.64	1.570	0.32	6.98	8.30	23.	2	
r	Promedio_tiempo_distribuci	583	65	061	261	906	848	3	0.000
1	ón_información_POST								

Pvalor=0.000

Pvalor<0.05 se rechaza la hipótesis nula.

Entonces se rechaza la hipótesis Ho y se acepta la hipótesis Ha, por lo tanto, se afirma que: Hay una diferencia significativa en las medias del

proceso de distribución de información antes y después de la implementación del sistema basado en Web Scraping en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja. Por lo cual se concluye que el uso de un sistema basado en Web Scraping influye positivamente en la distribución de información.

4.2.3. Prueba de la Hipótesis Específica N° 03

La implementación de un sistema basado en Web Scraping influye positivamente en la toma de decisiones en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

4.2.3.1. Planteamiento de las Hipótesis

Ho: El uso de un sistema basado en Web Scraping NO influye positivamente en la toma de decisiones en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja

Ha: El uso de un sistema basado en Web Scraping influye positivamente en la toma de decisiones en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja

4.2.3.2. Determinación del Nivel de Significancia

- Alfa = 95%.
- Error= 0.05

4.2.3.3. Elección de la Prueba Estadística

Se elige la prueba de T Student, para muestras relacionadas, por ser un estudio longitudinal con medidas del antes y después.

4.2.3.4. Cálculo del Valor Tabular:

NORMALIDAD:

Se elige la prueba de Shapiro Wilk (<33), porque nuestra muestra es menor a 33.

Criterios para determinar la Normalidad:

P valor $\Rightarrow \alpha$ Aceptar H_0 = Los datos provienen de una distribución normal

P valor $< \alpha$ Aceptar H_1 = Los datos NO provienen de una distribución normal

Tabla 22

Pruebas de normalidad Kolmogorov-Smirnov - Shapiro-Wilk

Pruebas de normalidad						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístic	gl	Sig.	Estadístic	gl	Sig.
	o			o		
Toma_decisiones_PRE	0.167	24	0.083	0.933	24	0.113
Toma_decisiones_POS	0.163	24	0.099	0.920	24	0.059
T						

a. Corrección de significación de Lilliefors

Se observa el P valor:

Tabla 23*Normalidad de Pvalor*

NORMALIDAD		
P valor (PRE PRUEBA) = 0.113	>	$\alpha = 0.05$
P-valor (POST PRUEBA) = 0.059	>	$\alpha = 0.05$
Conclusión: Los datos de la dimensión adquisición de Información provienen de una distribución normal.		

Decisión Estadística

Tabla 24*Estadísticas de muestras emparejadas*

Estadísticas de muestras emparejadas					
		Media	N	Desv. Desviación	Desv. Error promedio
Par 1	Toma_decisiones_PRE	1.7500	24	0.31277	0.06384
	Toma_decisiones_POST	4.0000	24	0.31277	0.06384

Fuente: Elaboración Propia**Tabla 25***Correlaciones de muestras emparejadas*

Correlaciones de muestras emparejadas	
--	--

	N	Correlación	Sig.
Par 1 & Toma_decisiones_PRE & Toma_decisiones_POST	24	-0.361	0.083

Fuente: Elaboración Propia

Tabla 26

Prueba de muestras emparejadas

		Diferencias emparejadas							
		Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	Sig. (bilateral)	
					Inferior	Superior			
Par 1	Toma_decisiones _PRE - Toma_decisiones _POST	2.2500	0.5160	0.10534	2.46791	2.03209	21.360	0.000	

Pvalor=0.000

Pvalor<0.05 se rechaza la hipótesis nula.

Entonces se rechaza la hipótesis Ho y se acepta la hipótesis Ha, por lo tanto, se afirma que: Hay una diferencia significativa en las medias del

proceso de toma de decisiones antes y después de la implementación del sistema basado en Web Scraping en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja. Por lo cual se concluye que El uso de un sistema basado en Web Scraping influye positivamente en la toma de decisiones.

4.3. Discusión de Resultados

4.3.1. *En relación a la variable independiente modelo de sistema basado en Web Scraping*

Durante la evaluación del modelo de sistema basado en Web Scraping para los indicadores funcionabilidad, usabilidad y satisfacción se mostró resultados de los cuales se detalla a continuación:

En el indicador funcionabilidad se obtiene un resultado de 16,5 sobre un total de 24 usuarios, con respecto a la evaluación del indicador usabilidad la calificación fue 18,5 de un total de 24 usuarios y por último del indicador satisfacción el resultado fue 21,5 de un total de 24 usuarios. Ello demuestra, la calidad de uso del modelo de sistema basado en Web Scraping según la ISO 25010/9126.

Estos resultados a su vez podemos contrastar con la investigación realizada por Belito y Mayon (2019) la evaluación de usabilidad de un sistema de prescripción médica tuvo un resultado de 25,5; del indicador satisfacción el resultado fue 30,5 y según Zuasnabar (2020) el indicador funcionalidad tuvo un resultado de 11 haciendo un 100% de los criterios evaluados.

4.3.2. *De acuerdo con la dimensión Adquisición de Información*

En una encuesta realizada a 24 personas de la UERST, se aplicó un cuestionario para medir el tiempo de adquisición de información, en la cual nos muestra que el

tiempo promedio de adquisición de información en la PRE TEST es de 21.3 días y el tiempo promedio de adquisición de información en la POST TEST es de 2.7 días.

De acuerdo al análisis de resultados obtenidos se demuestra que la implementación de un sistema basado en Web Scraping influye positivamente en la adquisición de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

Huamán & Quispe (2019) en su tesis determinan que el Web Scraping es una técnica que facilita la extracción de información de forma masiva y automatizada, para que pueda ser usada de acuerdo a las necesidades del usuario.

4.3.3. De acuerdo con la dimensión Distribución de Información

En una encuesta realizada a 24 personas de la UERST, se aplicó un cuestionario para medir el tiempo de distribución de información, en la cual nos muestra que el tiempo promedio de distribución de información en la PRE TEST es de 9.7 y el tiempo promedio de distribución de información en la POST TEST es de 2.08 días.

De acuerdo al análisis de resultados obtenidos se demuestra que la implementación de un sistema basado en Web Scraping influye positivamente en la distribución de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

Capuñay (2018) sostiene que se logró con éxito identificar la fuente de datos para el análisis del perfil laboral de los profesionales de ingeniería de sistemas para desempeñar cargos en instituciones públicas, así mismo, realizó con éxito la recolección de datos a partir de la aplicación de técnicas de Scraping de datos.

4.3.4. De acuerdo con la dimensión Toma de Decisiones

En una encuesta realizada a 24 personas de la UERST, se aplicó un cuestionario para medir el nivel en la toma de decisiones, en la cual nos muestra que el promedio de

toma de decisiones en la PRE TEST es de 1.75 quiere decir que están en desacuerdo con la toma de decisiones y el promedio en la POST TEST es de 4.0, quiere decir que después de la implementación del sistema Web Scraping están muy de acuerdo con la toma de decisiones en la UERST.

De acuerdo al análisis de resultados obtenidos se demuestra que la implementación de un sistema basado en Web Scraping influye positivamente en la toma de decisiones en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

Loayza (2015) sostiene que sí existe relación entre la Gestión de la Información y la Toma de Decisiones en la Dirección de Infraestructura de la Universidad Nacional José María Arguedas Andahuaylas, 2014. Concluye que la relación determinada es positiva media entre los variables en estudio.

CONCLUSIONES

1. Se concluye que, la implementación de un modelo de sistema basado en Web Scraping influye positivamente en la gestión de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja, siendo así que, en el indicador funcionalidad se obtiene un resultado de 16,5 sobre un total de 24 usuarios, con respecto a la evaluación del indicador usabilidad la calificación fue 18,5 de un total de 24 usuarios y por último del indicador satisfacción el resultado fue 21,5 de un total de 24 usuarios. Ello demuestra, la calidad de uso del modelo de sistema basado en Web Scraping según la ISO 25010/9126.
2. Se concluye que, La implementación de un sistema basado en Web Scraping influye positivamente en la Adquisición de Información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja, al evidenciar en la encuesta realizada a 24 personas, en la cual nos muestra que el tiempo promedio de adquisición de información se redujo de 21.3 días a 2.7 días. y al observar el valor de la prueba de hipótesis específica 01 , donde $P\text{valor} = 0.000 < 0.05$. Entonces se rechaza la hipótesis H_0 y se acepta la hipótesis H_a , en este sentido se afirma que: Hay una diferencia significativa en las medias del proceso de adquisición de información antes y después de la implementación del sistema basado en Web Scraping en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.
3. Se concluye que, La implementación de un sistema basado en Web Scraping influye positivamente en la Distribución de Información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja, al evidenciar en la encuesta realizada a 24 personas, en la cual nos muestra que el tiempo promedio de distribución de información se redujo de 9.7 días a 2.08 días. y al observar el valor de la prueba de hipótesis específica 01 , donde $P\text{valor} = 0.000 < 0.05$. Entonces se rechaza la hipótesis H_0 y se

acepta la hipótesis H_a , en este sentido se afirma que: Hay una diferencia significativa en las medias del proceso de distribución de información antes y después de la implementación del sistema basado en Web Scraping en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

4. Se concluye que, La implementación de un sistema basado en Web Scraping influye positivamente en la Toma de Decisiones en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja, al evidenciar en la encuesta realizada a 24 personas, en la cual nos muestra que que el promedio de toma de decisiones en la PRE TEST es de 1.75 quiere decir que están en desacuerdo con la toma de decisiones y el promedio en la POST TEST es de 4.0, quiere decir que están muy de acuerdo con la toma de decisiones después de la implementación del sistema Web Scraping. Y al observar el valor de la prueba de hipótesis específica 01, donde $P_{valor} = 0.000 < 0.05$. Entonces se rechaza la hipótesis H_0 y se acepta la hipótesis H_a , en este sentido se afirma que: Hay una diferencia significativa en las medias del nivel de toma de decisiones antes y después de la implementación del sistema basado en Web Scraping en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.

RECOMENDACIONES

1. Diseñar e implementar en un solo software para automatizar los procesos realizados en el presente proyecto de tesis (Web Scraping, SQL Server y reportes en Excel).
2. Capacitar constantemente a los responsables del área de informática y estadística para ampliar sus conocimientos en procesamiento de información de base de datos en SQL Server y los tableros en el Excel.
3. Destinar mayor presupuesto para la implementación de sistemas de información y para mejorar en el área de tecnologías de información y comunicación en la Unidad Ejecutora 403 Red de Salud Tayacaja.
4. Capacitar a los profesionales de la salud en la utilización de los tableros estadísticos en Excel.

REFERENCIA BIBLIOGRÁFICA

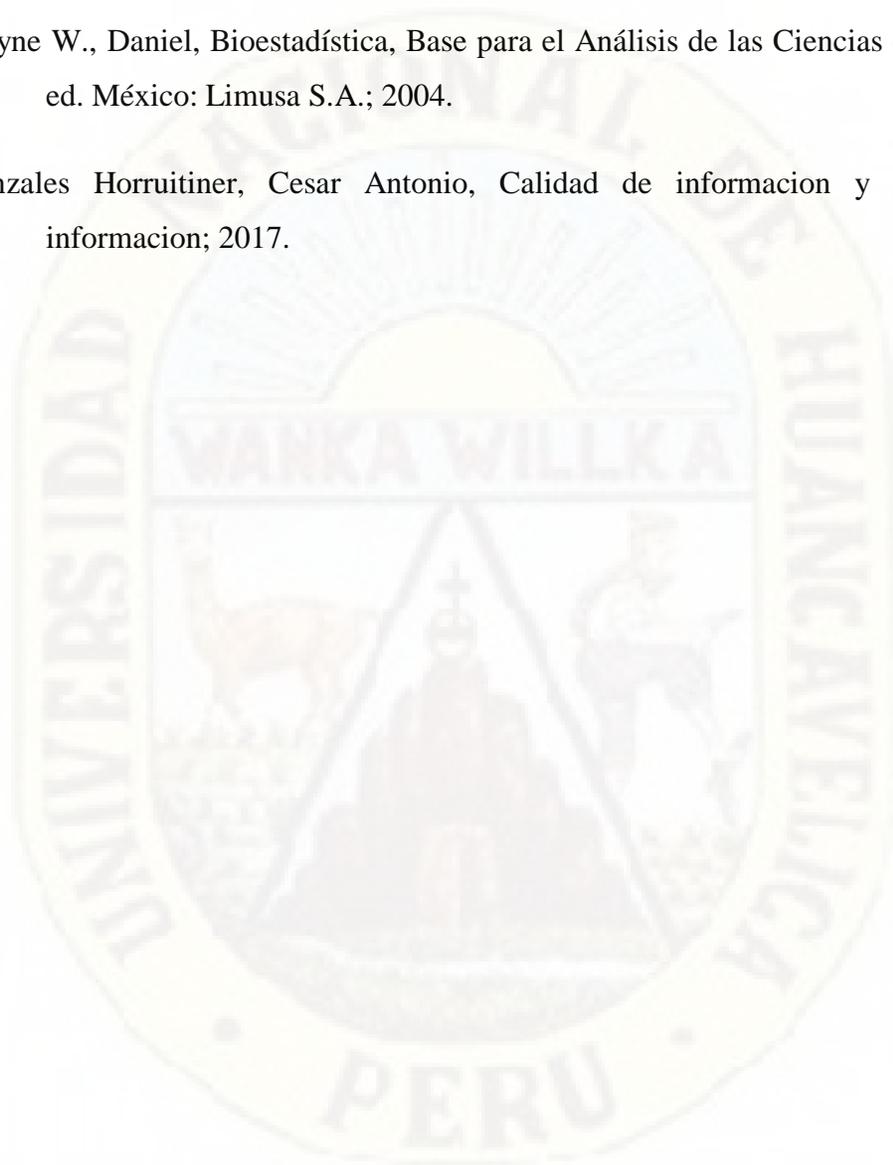
- Date, Cow: "Introducción a los sistemas de bases de datos". 1er. Edición Pearson Educación, 2001. ISBN 968-444-419.
- Domínguez Bravo, Javier. "Breve Introducción a la Cartografía y a los Sistemas de Información (SISTEMA BASADO EN WEB SCRAPING)". 1a ed. Madrid: CIEMAT; 2000.
- Olaya, Víctor. "Sistemas de Información" Versión 1.0. [sede Web]. 2010, disponible en: http://wiki.osgeo.org/wiki/Libro_SISTEMA_BASADO_EN_WEB_SCRAPING
- Espinoza, Ciro. "Metodología de investigación tecnológica". Pensando en sistemas. 1ra edición. Huancayo, Perú: Imagen Gráfica, marzo de 2010.
- Lardent, Alberto, R. "Sistemas de información para la gestión empresarial: planeamiento, tecnología y calidad". 1era. Edición. Editorial. Pearson. Buenos Aires. 2001.
- Ralph Kimball. 1996. "The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses". 3era edición. John Wiley & Sons . ISBN 0-471-15337-0.
- Chiaveanato, I; Administración de Recursos Humanos, 5a ed. Colombia: McGraw-Hill Interamericana; 2000.
- Chiaveanato, I; Introducción a la Teoría General de la Administración, 5a ed. México: McGraw-Hill Interamericana; 2005.
- Douglas C., Montgomery y George C. Runger, Probabilidad y Estadística Aplicadas a La Ingeniería. 2a ed. México: LIMUSA S.A.; 2004.

- Hernández Sampieri, Roberto; Metodología de la Investigación, 1a ed. México: McGraw-Hill Interamericana; 1999.
- Laudon, Kenneth C., Laudon, Jane P., Administración de los Sistemas de Información Organización y tecnología, 10a ed. México: Prentice Hall Hispanoamericano S.A.; 2008.
- Monzón F., J. & Spencer, David, Análisis y Diseño de Sistemas. 1a ed. Lima: Gómez; 1994.
- Munrray R., Spiegel, Shiller, Jhon y Alu Srinivasa R., Probabilidad y Estadística. 2a ed. México: McGraw-Hill; 2003.
- Pressman, Roger S. Ingeniería de Software un enfoque práctico, 3a ed. España: McGraw – Hill; 1998.
- Senn, James A. Análisis y Diseño de Sistemas de Información. 2a ed. México: McGraw-Hill; 1992.
- Star J. y Estes J. Geographic Information Systems: an Introduction. 1a ed. Madrid: Prentice-Hall; 1990.
- Tomlin, C.D. Geographic information systems and cartographic modeling. 1a ed. Madrid: Prentice Hall; 1990.
- Pando Fernández, Yuri J. SQL Server 2008, Diseño de Base de datos. 1a ed. Lima: Gómez; Marzo 2009.
- Vicente González José Luis, Behm Chang Virginia. “Consulta, Edición y Análisis Espacial con Arcgis 9.2. Tomo I”. [sede Web]. Junta De Castilla y León: LVG.pdf.2008. disponible en:

http://www.gabrielortiz.com/descargas/2008_Manual_TeoriaárcGIS92_VBCyJLVG.pdf
. 2008.

Wayne W., Daniel, Bioestadística, Base para el Análisis de las Ciencias de la Salud. 4a ed. México: Limusa S.A.; 2004.

Gonzales Horruitiner, Cesar Antonio, Calidad de informacion y gestion de la informacion; 2017.



INFORMES, ARTÍCULOS Y REVISTAS

Andrea Zvenger Patricia, “Introducción al soporte de decisiones”. Incorporación de soluciones OLAP en entornos empresariales. Tesis de licenciatura, Argentina; 2005.

D’hers Thierry – Microsoft Corporation & VICKERY Stacy – Progressive Strategies, “Comparing business intelligence platfoms” .USA; 2006

Ferrando Giron Santiago Jose&CarreiraPerpiñanIvan José. “Solucion datawarehouse bajo el nuevo modelo de organización e- business: la necesidad de integrar la información” telefónica investigación y desarrollo.

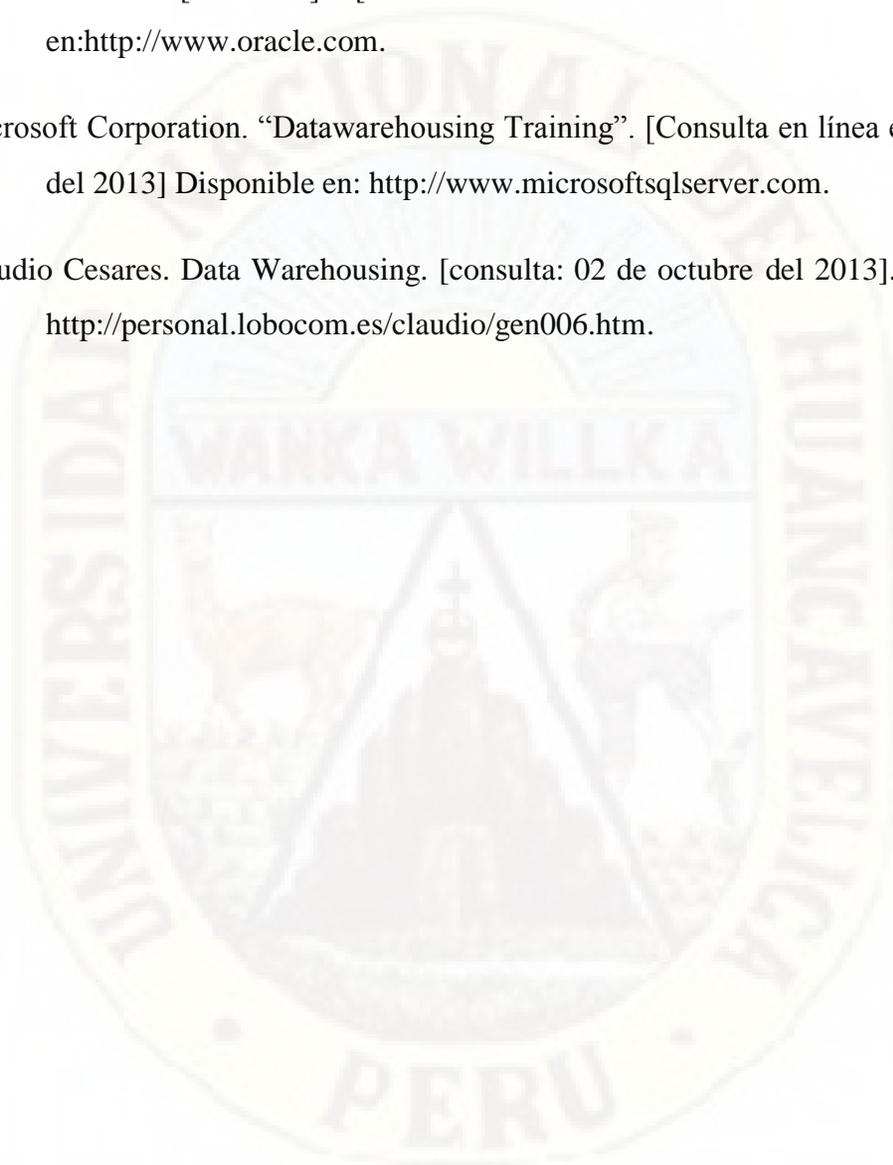
Mantero María Eugenia – Analista de DataWarehousing – RmyA S.R.L. “Diseño De Datawarehouse Performance”. Argentina; 2002.

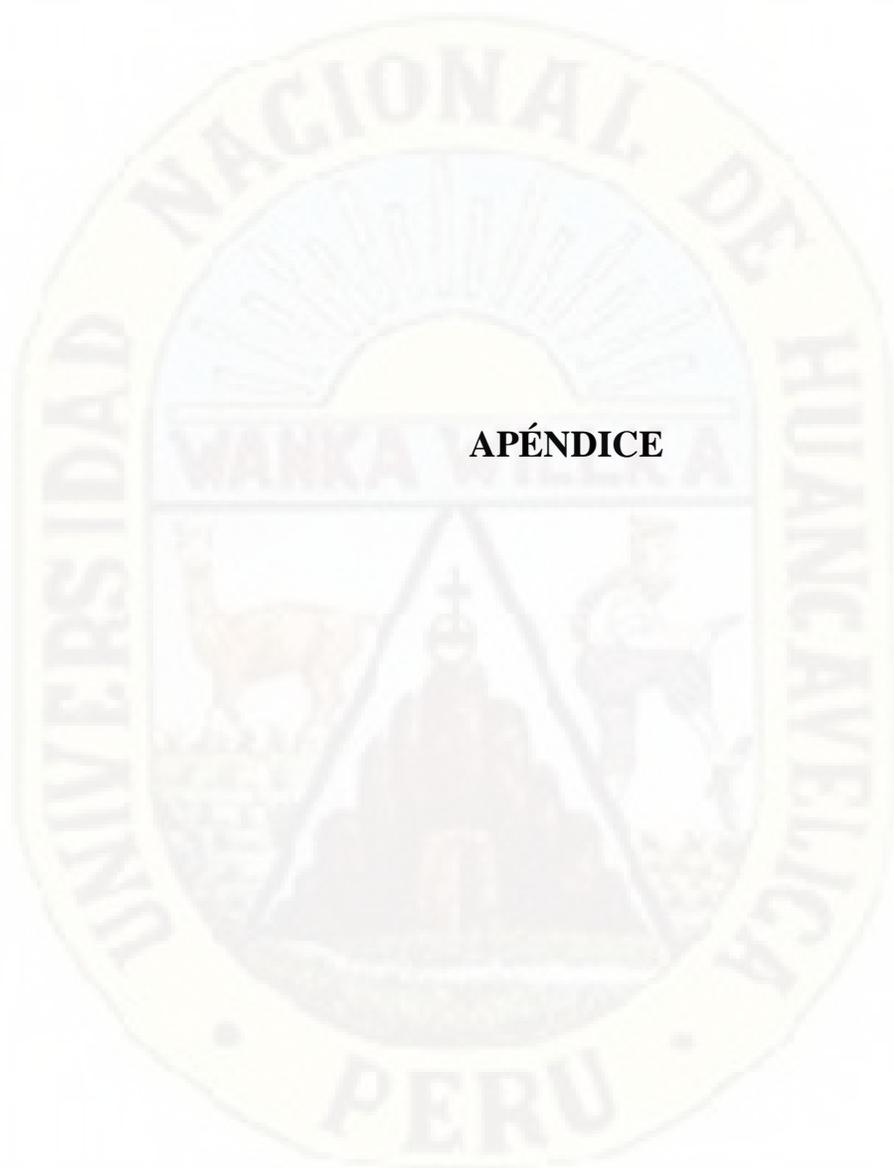
REFERENCIAS ELECTRÓNICAS

BI Solutions”. [en línea]. [Consulta: 26 de setiembre de 2012]. Disponible en:<http://www.oracle.com>.

Microsoft Corporation. “Datawarehousing Training”. [Consulta en línea el 27 setiembre del 2013] Disponible en: <http://www.microsoftsqlserver.com>.

Claudio Cesares. Data Warehousing. [consulta: 02 de octubre del 2013]. Disponible en <http://personal.lobocom.es/claudio/gen006.htm>.





APÉNDICE

MATRIZ DE CONSISTENCIA

Anexo 1. Matriz de consistencia

Título: MODELO DE SISTEMA BASADO EN WEB SCRAPING EN LA GESTIÓN DE INFORMACION EN LA UNIDAD DE SEGUROS PÚBLICOS Y PRIVADOS DE LA UNIDAD EJECUTORA 403 RED DE SALUD TAYACAJA 2019

PROBLEMA GENERAL	OBJETIVO GENERAL	HIPÓTESIS GENERAL	VARIABLE INDEPENDIENTE	VARIABLE DEPENDIENTES	METODOLOGÍA
¿Cómo influye la implementación de un modelo de sistema basado en Web Scraping en la Gestión de Información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja?	Demostrar cómo influye la implementación de un modelo de sistema basado en Web Scraping en la Gestión de Información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.	La implementación de un modelo de sistema basado en Web Scraping influye positivamente en la Gestión de Información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.	X. Sistema basado en Web Scraping	Y. Gestión de Información	<p>Tipo de Investigación:</p> <p>La investigación se enmarca dentro del tipo de investigación aplicada.</p> <p>Nivel de Investigación:</p> <p>Y el nivel de investigación es explicativo.</p> <p>Diseño de investigación:</p> <p>Diseño</p>
Problemas específicos	Objetivos específicos	Hipótesis específicas	Dimensiones	Dimensiones	n:

<p>¿Cómo influye la implementación de un modelo de sistema basado en Web Scraping en la adquisición de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja?</p>	<p>Demostrar cómo influye la implementación de un modelo de sistema basado en Web Scraping en la adquisición de Información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.</p>	<p>La implementación de un modelo de sistema basado en Web Scraping influye positivamente en la adquisición de Información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.</p>	<p>X1. Calidad de uso</p>	<p>Y1. Adquisición de Información</p>	<p>Pre - experimental I con un solo grupo donde se mide con un pre-test y post-test</p> <p>G: 0₁ X 0₂</p> <p>Dónde: G: Grupo experimental 0₁: Pre-test X: Sistema basado en Web Scraping 0₂: Post-test</p>
<p>¿Cómo influye la implementación de un sistema basado en Web Scraping en la distribución de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja?</p>	<p>Demostrar cómo influye la implementación de un sistema basado en Web Scraping en la distribución de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.</p>	<p>La implementación de un sistema basado en Web Scraping influye positivamente en la distribución de información en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.</p>	<p>Y2. Distribución de información</p> <p>Y3. Toma de decisiones</p>	<p>Y2. Distribución de información</p> <p>Y3. Toma de decisiones</p>	<p>Población y Muestra:</p> <p>24 Profesionales encargados de las áreas y estrategias sanitarias de la USPP de la UERST.</p> <p>TECNICA MUESTRA:</p> <p>L: ALEATORIO SIMPLE</p>

Red de Salud Tayacaja? ¿Cómo influye la implementación de un modelo de sistema basado en Web Scraping en la toma de decisiones en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja?	Ejecutora 403 Red de Salud Tayacaja. Demostrar cómo influye la implementación de un modelo de sistema basado en Web Scraping en la toma de decisiones en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.	Ejecutora 403 Red de Salud Tayacaja. La implementación de un modelo de sistema basado en Web Scraping influye positivamente en la toma de decisiones en la Unidad de Seguros Públicos y Privados de la Unidad Ejecutora 403 Red de Salud Tayacaja.	Muestra: 24 Profesionales encargados de las áreas y estrategias sanitarias de la USPP de la UERST.
---	--	---	--

Tabla 27

Definición operacional de las variables

Variable Independiente	Definición	Dimensiones	Indicadores
Modelo de sistema basado en Web Scraping	Modelo de sistema basado en Web Scraping, es un método utilizado para recopilar información de toda la web, el mismo que permite la satisfacción del usuario y mostrar un sistema adecuado de calidad.	Calidad de uso	<ul style="list-style-type: none"> • Funcionalidad • Usabilidad • Satisfacción
Variable dependiente	Definición	Dimensiones	Indicadores
Gestión de información	La gestión de la información es considerada como el proceso para la obtención de la información adecuada, en la forma correcta, para la persona u organización indicada, al precio adecuado, en el tiempo oportuno y lugar apropiado, para tomar la decisión adecuada. (Woodman, 1985).	Adquisición de información Distribución de información Toma de decisiones	<ul style="list-style-type: none"> • Tiempo de adquisición de datos (días) • Tiempo de procesamiento de datos (días) • Tiempo de distribución de información (días) • Tiempo de impresión de reportes (días) • Tipos • Condiciones • Modelo • Estilos

Matriz de instrumentos

Cuestionario pre test

Instrucciones: De los supuestos o de las situaciones que aquí se presentan, marque con un aspa (x) aquel indicador con el que usted se identifica más.

Variable independiente: Sin el modelo de sistema basado en Web Scraping - PRE TEST

Tabla 28

Cuestionario PRE TEST

Preguntas (Item)	Escala de valoración				
	En desacuerdo (1)	Poco de acuerdo (2)	De acuerdo (3)	Muy de acuerdo (4)	Totalmente de acuerdo (5)
D3: Toma de decisiones					
1. ¿La toma de decisiones, está en función del tipo de la información que se utiliza en la oficina?	1	2	3	4	5
2. ¿Dispones con información completa y confiable, seguro de lo que ocurrirá sobre la alternativa seleccionada, con pocas posibilidades de equivocarse?	1	2	3	4	5
3. ¿Realizas actividades con información completa y perfecta, eliminando la incertidumbre que maximice el logro de las metas de la oficina?	1	2	3	4	5
4. ¿Buscas información de	1	2	3	4	5

calidad, cantidad y momento oportuno para solucionar problemas complejos?

Cuestionario post test

Instrucciones: De los supuestos o de las situaciones que aquí se presentan, marque con un aspa (x) aquel indicador con el que usted se identifica más.

Variable Independiente: Con el modelo de sistema basado en Web Scraping - POST TEST

Tabla 29

Cuestionario POST TEST

Preguntas (Item)	Escala de valoración				
	En desacuerdo (1)	Poco de acuerdo (2)	De acuerdo (3)	Muy de acuerdo (4)	Totalmente de acuerdo (5)
D3: Toma de decisiones					
1. ¿La toma de decisiones, está en función del tipo de la información que se utiliza en la oficina, después de la implementación del sistema Web Scraping?	1	2	3	4	5
2. ¿Dispones con información completa y confiable, seguro de lo que ocurrirá sobre la alternativa seleccionada, con pocas posibilidades de equivocarse, después de la implementación del sistema Web Scraping?	1	2	3	4	5
3. ¿Realizas actividades con información completa y perfecta, eliminando la incertidumbre que maximice el logro de las metas de la oficina, después de la implementación del sistema Web	1	2	3	4	5

Scraping?

4. ¿Buscas información de calidad, cantidad y momento oportuno para solucionar problemas complejos, después de la implementación del sistema Web Scraping?

1 2 3 4 5

Modelo de sistema basado en Web Scraping

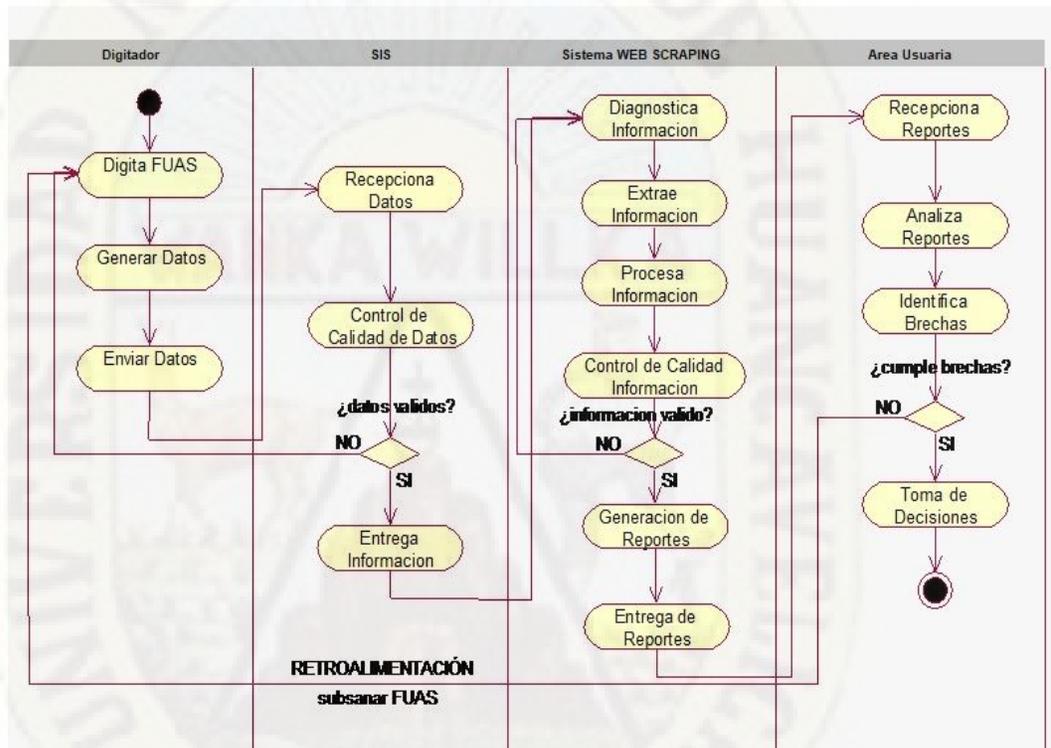


Figura 17. Modelo de sistema basado en Web Scraping

Fuente: Elaboración propia

Procesos de desarrollo de Web Scraping

Web Scraping es el proceso de recopilar información de forma automática de la Web. Es un campo con desarrollos activos, compartiendo un propósito en común con la visión de la Web semántica. Utiliza soluciones prácticas basadas en tecnologías existentes que son comúnmente ad hoc

Instalación

Puede instalar la extensión desde la tienda Chrome, después de instalarlo debe reiniciar Chrome para asegurarse de que la extensión esté completamente cargada. Si no desea reiniciar Chrome use la extensión solo en las pestañas que se crean después de instalarlo.

Requisitos

La extensión requiere Chrome 49+. No hay limitaciones del sistema operativo.

Open web scraper

Web scraper está integrado en las herramientas de desarrollador de Chrome, también puede usar estos accesos directos para abrir las herramientas de desarrollador, después de abrir herramientas para desarrolladores, abra la pestaña Web Scraper.

Shourtcuts:

Windows, Linux: Ctrl+Shift+I, f12, abierto Tools / Developer tools

Mac Cmd+Opt+I, abierto Tools / Developer tools

Configuración de extensiones Chrome

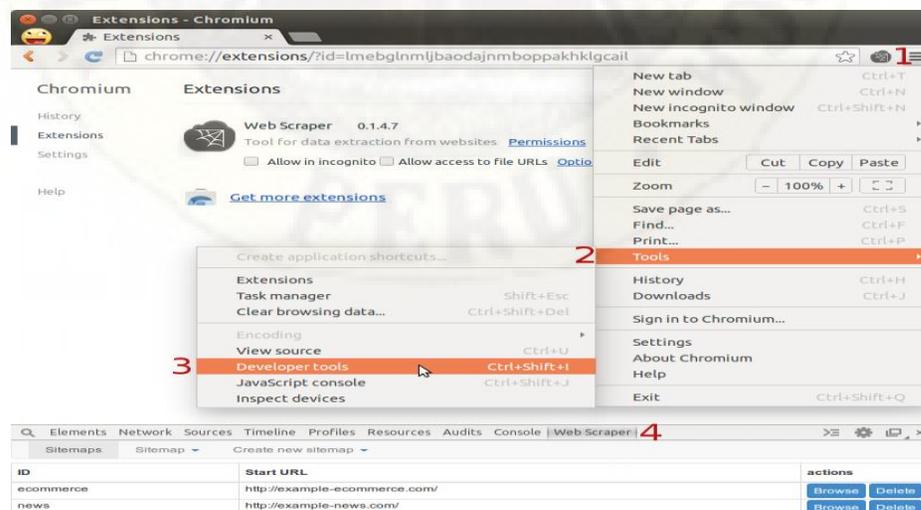


Figura 18. Configuración de extensiones

Raspando un sitio

Abra el sitio que desea raspar.

Crear mapa del sitio

Lo primero que debe hacer al crear un mapa del sitio es especificar la URL de inicio, esta es la URL desde la que comenzará el raspado; También puede especificar múltiples URL de inicio si el raspado debe comenzar desde múltiples lugares. Por ejemplo, si desea raspar varios resultados de búsqueda puede crear una URL de inicio separada para cada resultado de búsqueda.

Especificar múltiples URL con rangos

En los casos en que un sitio utiliza la numeración en las URL de las páginas, es mucho más sencillo crear una URL de inicio de rango que crear selectores de enlaces que naveguen por el sitio. Para especificar una URL de rango, reemplace la parte numérica de la URL de inicio con una definición de rango - [1-100]. Si el sitio utiliza cero rellenos en las URL, agregue cero rellenos a la definición del rango - [001-100]. Si desea omitir algunas URL, también puede especificar incremental como este [0-100:10].

Use la URL de rango como esta `http://example.com/page/ [1-3]` para enlaces como estos:

`http://example.com/page/1`

`http://example.com/page/2`

`http://example.com/page/3`

Use la URL de rango con cero rellenos como este `http://example.com/page/ [001-100]` para enlaces como estos:

`http://example.com/page/001`

`http://example.com/page/002`

`http://example.com/page/003`

Utilice la URL de rango con incrementos como este `http://example.com/page/ [0-100:10]` para enlaces como estos:

`http://example.com/page/0`

`http://example.com/page/10`

`http://example.com/page/20`

Crear selectores

Después de haber creado el mapa del sitio, puede agregarle selectores. En el panel selectores puede agregar nuevos selectores, modificarlos y navegar por el árbol de selectores. Los selectores se pueden agregar en una estructura de tipo árbol. El raspador web ejecutará los selectores en el orden en que están organizados en la estructura de árbol. Por ejemplo, hay un sitio de noticias y desea raspar todos los artículos cuyos enlaces están disponibles en la primera página. En la imagen 20 puede ver este sitio de ejemplo.

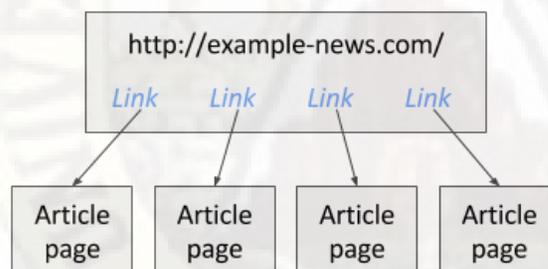


Figura 19. Ejemplo de selector de enlaces

Para raspar este sitio, puede crear un selector de enlace que extraerá todos los enlaces del artículo en la primera página. Luego, como selector secundario, puede agregar un selector de texto que extraerá artículos de las páginas de artículos a los que el selector de enlaces encontró enlaces. La imagen a continuación ilustra cómo se debe construir el mapa del sitio para el sitio de noticias.

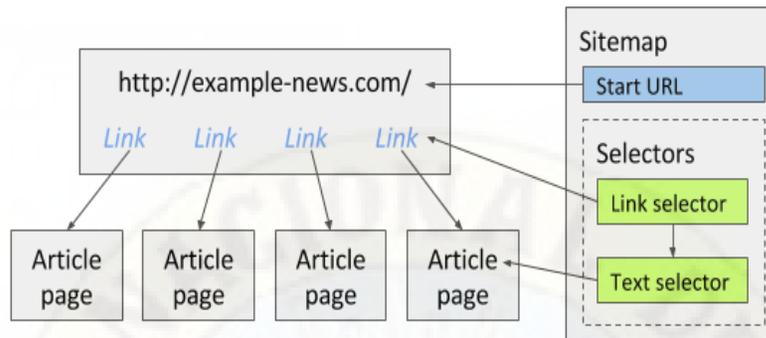


Figura 20. Vista previa de los selectores de enlaces

Tenga en cuenta que al crear selectores use las funciones vista previa de elementos y vista previa de datos para asegurarse de que ha seleccionado los elementos correctos con los datos correctos.

Más información sobre la construcción del árbol de selectores está disponible en la documentación del selector. Al menos debería leer sobre estos selectores principales:

- Selector de texto
- Selector de enlace
- Selector de elementos
- Inspeccionar el árbol selector

Después de haber creado los selectores para el mapa del sitio, puede inspeccionar la estructura de árbol de los selectores en el panel de gráficos del Selector. La imagen a continuación muestra un ejemplo de gráfico selector.



Figura 21. Ejemplo del grafico selector

Raspe el sitio

Después de haber creado selectores para el mapa del sitio, puede comenzar a raspar. Abra el panel raspar y comience a raspar. Se abrirá una nueva ventana emergente en la que el raspador cargará páginas y extraerá datos de ellas. Una vez realizado el raspado, la ventana emergente se cerrará y se le notificará con un mensaje emergente. Puede ver los datos raspados abriendo el panel examinar y exportarlos abriendo exportar datos como panel CSV.

Selectores

Web scraper tiene múltiples selectores que pueden usarse para diferentes tipos de extracción de datos y para diferentes interacciones con el sitio web. Los selectores se pueden dividir en tres grupos:

- Selectores de extracción de datos para extracción de datos.
- Selectores de enlaces para la navegación del sitio.
- Selectores de elementos para la selección de elementos que separan múltiples registros

Selectores de extracción de datos

Los selectores de extracción de datos simplemente devuelven datos del elemento seleccionado. Por ejemplo, el selector de texto extrae el texto del elemento seleccionado. Estos selectores pueden usarse como selectores de extracción de datos:

Selector de texto

El selector de texto se usa para la selección de texto. El selector de texto extraerá texto del elemento seleccionado y de todos sus elementos secundarios. HTML se eliminará y solo se devolverá el texto. El selector ignorará el texto dentro `<script>` y las `<style>` etiquetas. Las nuevas `
` etiquetas de línea serán reemplazadas por caracteres de nueva línea. También puede aplicar una expresión regular a los datos resultantes.

Opciones de configuración

Selector: selector CSS para el elemento del que se extraerán los datos.

Múltiple: se extraen varios registros. Por lo general, no se debe verificar. Si desea usar múltiples selectores de texto dentro de una página con múltiples marcados, entonces es posible que necesite un selector de elementos.

regex: expresión regular para extraer una subcadena del resultado.

Regex

El atributo de expresión regular se puede usar para extraer una subcadena del texto que extrae el selector. Cuando se utiliza una expresión regular, se devuelve toda la coincidencia (grupo 0) como resultado. www.regexr.com es un gran sitio donde puede aprender sobre expresiones regulares y probarlas.

Aquí hay algunos ejemplos que pueden resultarle útiles:

Tabla 30
Opciones de configuración

Texto	regex	Resultado
precio: 14.99 \$	[0-9]+\.[0-9]+	14,99
id: H83JKDX4	[A-Z0-9]{8}	H83JKDX4
fecha: 2014-08-20	[0-9]{4}\-[0-9]{2}\-[0-9]{2}	2014-08-20

Extraiga un registro por página con múltiples selectores de texto

Por ejemplo, está raspando un sitio de noticias que tiene un artículo por página. La página puede contener el artículo, su título, fecha de publicación y el autor. Un selector de enlaces puede navegar el raspador a cada una de estas páginas de artículos. Múltiples selectores de texto pueden extraer el título, la fecha, el autor y el artículo. Las opciones múltiples deben dejarse sin marcar para los selectores de texto porque cada página extrae solo un registro.

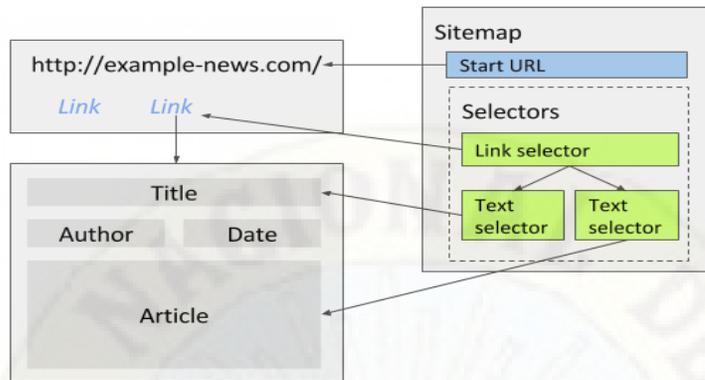


Figura 22. Configuración de selector de enlaces

Extraiga múltiples elementos con múltiples selectores de texto por página

Los sitios de comercio electrónico generalmente tienen varios artículos por página. Si desea raspar estos elementos, necesitará un selector de elementos que seleccione elementos de contenedor de elementos y múltiples selectores de texto que seleccionen datos dentro de cada elemento de contenedor de elementos.

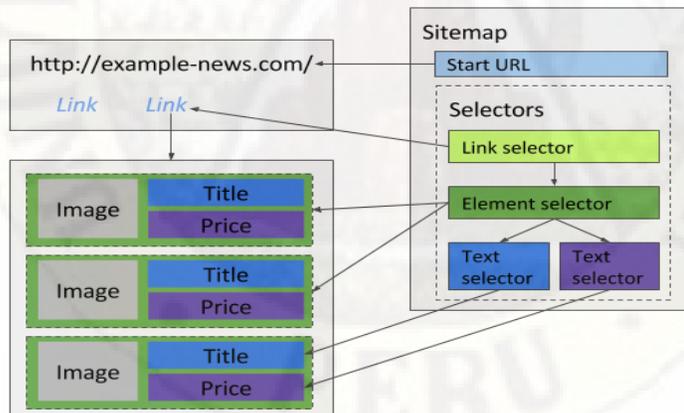


Figura 23. Selección de múltiples selectores de enlaces

Extraer múltiples registros de texto por página

Por ejemplo, desea extraer comentarios para un artículo. Hay varios comentarios en una sola página y solo necesita el texto del comentario (si necesita otros atributos de comentarios, consulte el ejemplo anterior). Puede usar el selector de texto para extraer

estos comentarios. El atributo múltiple de selectores de texto debe verificarse porque extraerá varios registros.

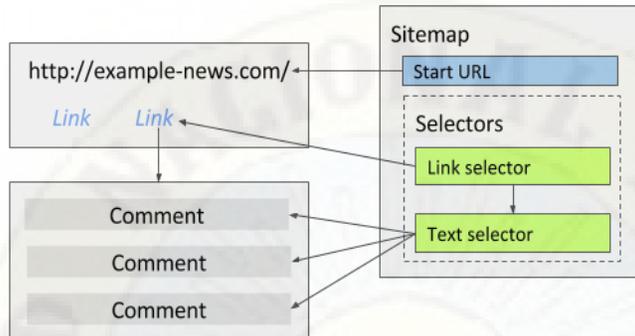


Figura 24. Selectores de texto

- Selector de enlace
- Selector emergente de enlace
- Selector de imagen
- Selector de mesa
- Selector de atributo de elemento
- Selector HTML
- Selector agrupado

Selectores de enlaces

Los selectores de enlaces extraen las URL de los enlaces que luego se pueden abrir para la extracción de datos. Por ejemplo, si en un árbol de mapa del sitio hay un selector de enlaces que tiene 3 selectores de texto secundarios, Web Scraper extrae todas las URL con el selector de enlaces y luego abre cada enlace y usa esos selectores de extracción de datos secundarios para extraer datos. Por supuesto, un selector de enlace podría tener selectores de enlace como selectores secundarios, entonces estos selectores de enlace secundarios se usarían para navegar más en la página. Estos son actualmente los selectores de enlaces disponibles:

Selector de enlace

Selector emergente de enlace

Selectores de elementos

Los selectores de elementos son para la selección de elementos que contienen múltiples elementos de datos. Por ejemplo, un selector de elementos podría usarse para seleccionar una lista de artículos en un sitio de comercio electrónico. El selector devolverá cada elemento seleccionado como elemento primario a sus selectores secundarios. Los selectores secundarios extraerán datos solo dentro del elemento que el selector de elementos les dio. Estos son actualmente los selectores de elementos disponibles:

- Selector de elementos
- Selector de desplazamiento hacia abajo del elemento
- Selector de clic de elemento

Opciones de configuración del selector

Cada selector tiene opciones de configuración. Aquí puedes ver los más comunes. Las opciones de configuración que son específicas de un selector se describen en la documentación de los selectores.

selector: selector CSS que selecciona un elemento en el que trabajará el selector.

múltiple: debe verificarse cuando se extraigan varios registros (filas de datos) con este selector. Los datos extraídos de dos o más selectores con múltiples comprobaciones no se fusionarán en un solo registro.

delay: retraso antes de utilizar el selector.

Selectores principales: configure los selectores principales para este selector para hacer el árbol del selector.

¡Nota! Un error común cuando se usa la opción de configuración múltiple es crear dos selectores junto con múltiples seleccionados y esperar que el raspador unirá los valores del selector en pares. Por ejemplo, si seleccionó enlaces de paginación y enlaces de navegación, estos enlaces no podrían unirse lógicamente en pares. La forma correcta es seleccionar un elemento contenedor con el selector de elementos y agregar selectores

de datos como selectores secundarios al selector de elementos con la opción múltiple no marcada.

Selector de CSS

Web Scraper utiliza selectores CSS para buscar elementos HTML en páginas web y extraer datos de ellos. Al seleccionar un elemento, el Web Scraper intentará adivinar cuál podría ser el selector CSS para los elementos seleccionados. Pero también puede escribirlo usted mismo y probarlo haciendo clic en "Vista previa del elemento". Puede usar selectores CSS que están disponibles en las versiones 1-3 de CSS y también pseudo selectores que están disponibles adicionalmente en jQuery. Aquí hay algunos enlaces de documentación que pueden ayudarlo:

Selectores CSS

Selectores jQuery CSS

Referencia del selector de CSS de w3schools

Selectores adicionales de Web Scraper

Es posible agregar nuevos pseudo selectores CSS a Web Scraper. En este momento solo hay un selector CSS agregado.

Selector de padres

CSS Selector_parent_permite que un selector secundario de un selector de elementos seleccione el elemento que fue devuelto por el selector de elementos. Por ejemplo, este selector CSS podría usarse en un caso en el que necesite extraer un atributo del elemento que devolvió el selector Elemento.

Seleccionar elementos dentro de iframe o root shadow

Se puede usar un selector CSS escrito personalizado para seleccionar elementos dentro de un iframe o shadow-root.

Para seleccionar datos dentro del iframe, cree un selector que seleccione el elemento del iframe y luego agregue: `iframe`. Por ejemplo: `iframe: iframe. selector-within-iframe`

Del mismo modo, se puede utilizar un selector CSS personalizado para seleccionar elementos dentro de una raíz sombra. Primero seleccione el elemento primario raíz y luego agréguelo: `shadow-root` selector. Por ejemplo: `shadow-root-parent-element: shadow-root. selector-within-shadow-root`

Exportación de datos

La extensión del navegador Web Scraper admite la exportación de datos en formato CSV, mientras que Web Scraper Cloud admite la exportación de datos en formatos CSV, XLSX y JSON. Los formatos XLSX y JSON se agregarán a la extensión Web Scraper en una versión futura.

Exportación de datos desde la extensión Web Scraper

Descargue los datos raspados a través de la `Export data as CSV` selección del Sitemapmenú debajo del menú. Los datos también se pueden descargar mientras se ejecuta el raspador.

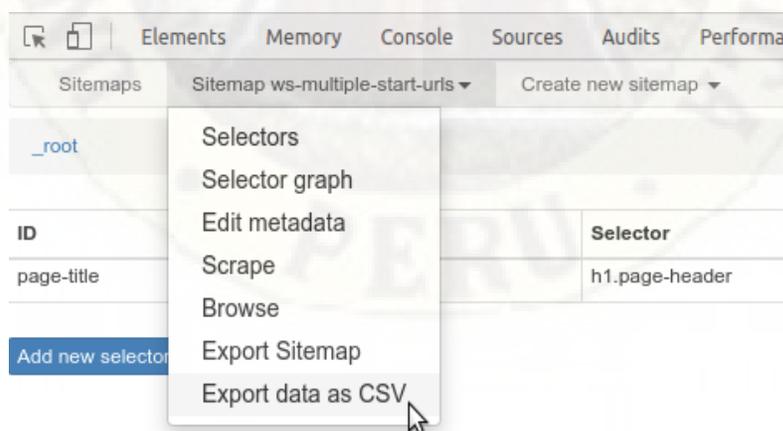


Figura 25. Exportación de data as CSV

Descargar datos de web scraper cloud

Descargue datos raspados a través del sitio web jobso de sistemas secciones. Los datos también se pueden descargar mientras se ejecuta el raspador.

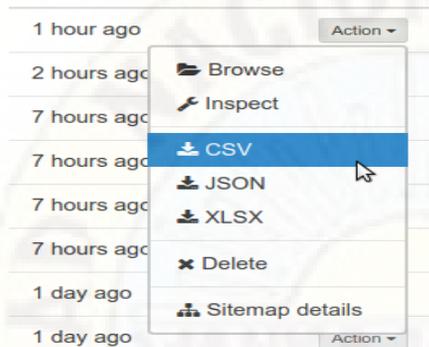


Figura 26. Exportación automatizada de datos

Exportación automatizada de datos

Configure la exportación de datos automatizada a Dropbox través de la Data Export sección. Los datos exportados actualmente estarán en formato CSV. Los datos se exportarán a Apps/Web Scraper su Dropbox carpeta.

Exportación de datos a través de API

Además, puede descargar datos a través de API de Web Scraper Cloud en formatos CSV o JSON.

Estructura del archivo de formato de datos y limitaciones

XLSX

Los datos en celdas separadas están limitados a 32767 caracteres. Personajes adicionales serán cortados. Utilice otros formatos de exportación si se esperan grandes contenidos de texto en una sola celda. El recuento de filas está limitado a 1 millón de filas. En caso de que el conjunto de datos contenga más de 1 millón de filas, los datos se dividirán en varias sub-hojas.

	A	B	C
1	title	price	description
2	Nokia 123	\$24.99	7 day battery
3	ProBook	\$739.99	14", Core i5 2.6GHz, 4GB, 500GB, Win7 Pro 64bit
4	ThinkPad X240	\$1311.99	12.5", Core i5-4300U, 8GB, 240GB SSD, Win7 Pro 64bit
5	Aspire E1-572G	\$581.99	15.6", Core i5-4200U, 8GB, 1TB, Radeon R7 M265, Windows 8.1
6	Lenovo IdeaTab	\$69.99	7" screen, Android
7	Acer Iconia	\$96.99	7" screen, Android, 16GB
8	Asus MeMO Pad	\$102.99	7" screen, Android, 8GB
9	Amazon Kindle	\$103.99	6" screen, wifi
10	iPad Mini Retina	\$537.99	Wi-Fi + Cellular, 32GB, Silver
11	IdeaTab A3500L	\$88.99	Black, 7" IPS, Quad-Core 1.2GHz, 8GB, Android 4.2
12	LG Optimus	\$57.99	3.2" screen
13	Samsung Galaxy	\$93.99	5 mpx. Android 5.0
14	Nokia X	\$109.99	Andoid, Jolla dualboot
15	Sony Xperia	\$118.99	GPS, waterproof
16	Ubuntu Edge	\$499.99	Sapphire glass
17	Packard 255 G2	\$416.99	15.6", AMD E2-3800 1.3GHz, 4GB, 500GB, Windows 8.1
18	Aspire E1-510	\$306.99	15.6", Pentium N3520 2.16GHz, 4GB, 500GB, Linux
19	ThinkPad T540p	\$1178.99	15.6", Core i5-4200M, 4GB, 500GB, Win7 Pro 64bit

Figura 27. Conjunto de datos

JSON

El formato de archivo JSON contiene un registro JSON por línea. Los nuevos caracteres de línea encontrados en los datos se escaparán, por "\n" lo que los \n caracteres se pueden usar de forma segura como separador de registros.

¡Nota! Analizar todo el archivo como una cadena JSON no funcionará, ya que todos los registros no están envueltos en una matriz JSON. Esta fue una decisión de diseño para facilitar el análisis de archivos grandes.

```
{ "title": "Nokia 123", "price": "$24.99", "description": "7 day battery" }
{ "title": "ProBook", "price": "$739.99", "description": "14\", Core i5 2.6GHz, 4GB, 500GB, Win7 Pro 64bit" }
{ "title": "ThinkPad X240", "price": "$1311.99", "description": "12.5\", Core i5-4300U, 8GB, 240GB SSD, Win7 Pro 64bit" }
{ "title": "Aspire E1-572G", "price": "$581.99", "description": "15.6\", Core i5-4200U, 8GB, 1TB, Radeon R7 M265, Windows 8.1" }
{ "title": "Lenovo IdeaTab", "price": "$69.99", "description": "7\" screen, Android" }
{ "title": "Acer Iconia", "price": "$96.99", "description": "7\" screen, Android, 16GB" }
{ "title": "Asus MeMO Pad", "price": "$102.99", "description": "7\" screen, Android, 8GB" }
{ "title": "Amazon Kindle", "price": "$103.99", "description": "6\" screen, wifi" }
{ "title": "iPad Mini Retina", "price": "$537.99", "description": "Wi-Fi + Cellular, 32GB, Silver" }
{ "title": "IdeaTab A3500L", "price": "$88.99", "description": "Black, 7\" IPS, Quad-Core 1.2GHz, 8GB, Android 4.2" }
{ "title": "LG Optimus", "price": "$57.99", "description": "3.2\" screen" }
{ "title": "Samsung Galaxy", "price": "$93.99", "description": "5 mpx. Android 5.0" }
{ "title": "Nokia X", "price": "$109.99", "description": "Andoid, Jolla dualboot" }
{ "title": "Sony Xperia", "price": "$118.99", "description": "GPS, waterproof" }
{ "title": "Ubuntu Edge", "price": "$499.99", "description": "Sapphire glass" }
{ "title": "Packard 255 G2", "price": "$416.99", "description": "15.6\", AMD E2-3800 1.3GHz, 4GB, 500GB, Windows 8.1" }
{ "title": "Aspire E1-510", "price": "$306.99", "description": "15.6\", Pentium N3520 2.16GHz, 4GB, 500GB, Linux" }
{ "title": "ThinkPad T540p", "price": "$1178.99", "description": "15.6\", Core i5-4200M, 4GB, 500GB, Win7 Pro 64bit" }
```

Figura 28. Registro de información JSON

CSV

Comma Separated Values El formato de los archivos se formatea como se describe en el estándar RFC 4180. Los valores se citan entre comillas dobles y en caso de que un carácter de comillas dobles esté en texto, se escapa con otro carácter de comillas dobles. Las líneas se separan con CR+LF \r\n caracteres. Además, los archivos CSV incluyen caracteres de marca de orden de bytes (BOM) U+FEFF al comienzo del archivo para indicar que el archivo estará en formato UTF-8. No se escapan los nuevos caracteres de línea, lo que significa que el uso \r\n como separador de registros puede generar errores. Recomendamos utilizar una biblioteca de lector CSV al leer archivos CSV mediante programación.

```
title,price,description
"Nokia 123","$24.99","7 day battery"
"ProBook","$739.99","14\"", Core i5 2.6GHz, 4GB, 500GB, Win7 Pro 64bit"
"ThinkPad X240","$1311.99","12.5\"", Core i5-4300U, 8GB, 240GB SSD, Win7 Pro 64bit"
"Aspire E1-572G","$581.99","15.6\"", Core i5-4200U, 8GB, 1TB, Radeon R7 M265, Windows 8.1"
"Lenovo IdeaTab","$69.99","7\"" screen, Android"
"Acer Iconia","$96.99","7\"" screen, Android, 16GB"
"Asus MeMO Pad","$102.99","7\"" screen, Android, 8GB"
"Amazon Kindle","$103.99","6\"" screen, wifi"
"iPad Mini Retina","$537.99","Wi-Fi + Cellular, 32GB, Silver"
"IdeaTab A3500L","$88.99","Black, 7\"" IPS, Quad-Core 1.2GHz, 8GB, Android 4.2"
"LG Optimus","$57.99","3.2\"" screen"
"Samsung Galaxy","$93.99","5 mp. Android 5.0"
"Nokia X","$109.99","Andoid, Jolla dualboot"
"Sony Xperia","$118.99","GPS, waterproof"
"Ubuntu Edge","$499.99","Sapphire glass"
"Packard 255 G2","$416.99","15.6\"", AMD E2-3800 1.3GHz, 4GB, 500GB, Windows 8.1"
"Aspire E1-510","$306.99","15.6\"", Pentium N3520 2.16GHz, 4GB, 500GB, Linux"
"ThinkPad T540p","$1178.99","15.6\"", Core i5-4200M, 4GB, 500GB, Win7 Pro 64bit"
```

Figura 29. Registro de información CSV

Abrir archivo CSV con un programa de hoja de cálculo

Recomendamos usar Libre Office Calc al abrir archivos CSV. Microsoft Office a menudo interpreta incorrectamente archivos CSV formateados en el estándar RFC 4180. Principalmente, esto está relacionado con el texto, incluidos los nuevos caracteres de línea.

En caso de que Microsoft Excel abra incorrectamente un archivo CSV, intente usar la función de importación de datos:

Comience con un archivo vacío

Ir a la pestaña Datos

Elija de texto / CSV

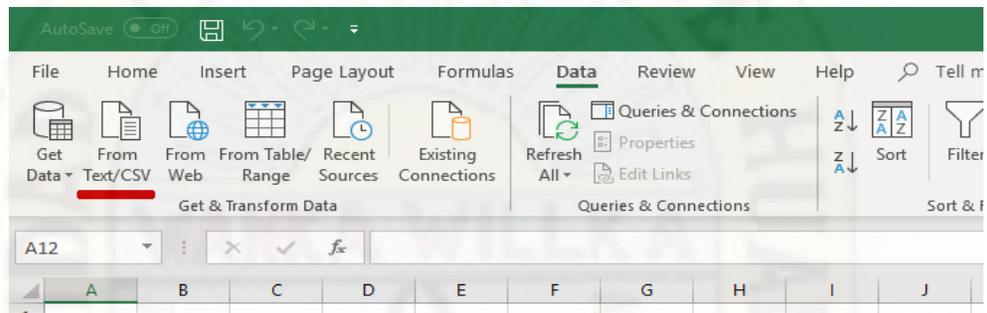


Figura 30. Pantalla Excel

Fuente: Elaboración propia

Abrir archivo CSV

Configurar ajustes de importación: codificación UTF-8, delimitador de coma, no detectar tipos de datos



Figura 31. Configuración de ajustes de importación

Cargar datos

Codificación de Web Scraping

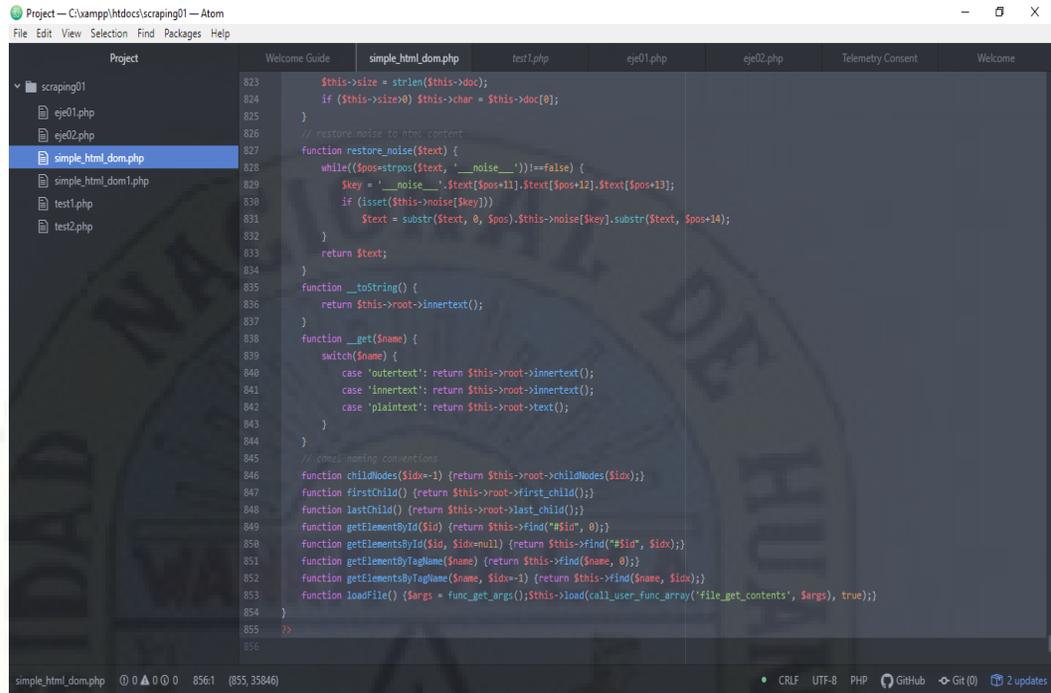


Figura 32. Codificación del Web Scraping

Fuente: Elaboración propia

<? php

```

/*****
*****/

```

Version: 1.11 (\$Rev: 175 \$)

```

*****/
*****/

```

```

define ('HDOM_TYPE_ELEMENT', 1);
define ('HDOM_TYPE_COMMENT', 2);
define ('HDOM_TYPE_TEXT', 3);
define ('HDOM_TYPE_ENDTAG', 4);
define ('HDOM_TYPE_ROOT', 5);
define ('HDOM_TYPE_UNKNOWN', 6);
define ('HDOM_QUOTE_DOUBLE', 0);
define ('HDOM_QUOTE_SINGLE', 1);
define ('HDOM_QUOTE_NO', 3);
define ('HDOM_INFO_BEGIN', 0);

```

```

define ('HDOM_INFO_END', 1);
define ('HDOM_INFO_QUOTE', 2);
define ('HDOM_INFO_SPACE', 3);
define ('HDOM_INFO_TEXT', 4);
define ('HDOM_INFO_INNER', 5);
define ('HDOM_INFO_OUTER', 6);
define('HDOM_INFO_ENDSPACE',7);

// helper functions
// -----
// get html dom from file
function file_get_html () {
    $dom = new simple_html_dom;
    $args = func_get_args ();
    $dom->load (call_user_func_array ('file_get_contents', $args), true);
    return $dom;
}

// get html dom from string
function str_get_html ($str, $lowercase=true) {
    $dom = new simple_html_dom;
    $dom->load ($str, $lowercase);
    return $dom;
}

// dump html dom tree
function dump_html_tree ($node, $show_attr=true, $deep=0) {
    $lead = str_repeat (' ', $deep);
    echo $lead. $node->tag;
    if ($show_attr && count($node->attr)>0) {
        echo '(';
        foreach($node->attr as $k=>$v)
            echo "[$k]=>\"". $node->$k."\" , ";
        echo ')';
    }
    echo "\n";
    foreach($node->nodes as $c)
        dump_html_tree ($c, $show_attr, $deep+1);
}

```

```

}
// get dom form file (deprecated)
function file_get_dom () {
    $dom = new simple_html_dom;
    $args = func_get_args ();
    $dom->load (call_user_func_array ('file_get_contents', $args), true);
    return $dom;
}
// get dom form string (deprecated)
function str_get_dom ($str, $lowercase=true) {
    $dom = new simple_html_dom;
    $dom->load ($str, $lowercase);
    return $dom;
}
// simple html dom node
// -----
class simple_html_dom_node {
    public $nodetype = HDOM_TYPE_TEXT;
    public $tag = 'text';
    public $attr = array ();
    public $children = array ();
    public $nodes = array ();
    public $parent = null;
    public $_ = array ();
    private $dom = null;
    function __construct($dom) {
        $this->dom = $dom;
        $dom->nodes [] = $this;
    }
    function __destruct () {
        $this->clear ();
    }
    function __toString () {
        return $this->outertext ();
    }
}

```

```

// clean up memory due to php5 circular references memory leak...
function clear () {
    $this->dom = null;
    $this->nodes = null;
    $this->parent = null;
    $this->children = null;
}
// dump node's tree
function dump($show_attr=true) {
    dump_html_tree ($this, $show_attr);
}
// returns the parent of node
function parent () {
    return $this->parent;
}
// returns children of node
function children($idx=-1) {
    if ($idx===-1) return $this->children;
    if (isset($this->children[$idx])) return $this->children[$idx];
    return null;
}
// returns the first child of node
function first_child () {
    if (count($this->children)>0) return $this->children [0];
    return null;
}
// returns the last child of node
function last_child () {
    if (($count=count($this->children))>0) return $this->children[$count-1];
    return null;
}
// returns the next sibling of node
function next_sibling () {
    if ($this->parent===null) return null;
    $idx = 0;

```

```

$count = count($this->parent->children);
while ($idx<$count && $this! ==$this->parent->children[$idx])
    ++$idx;
if (++$idx>=$count) return null;
return $this->parent->children[$idx];
}
// returns the previous sibling of node
function prev_sibling () {
    if ($this->parent===null) return null;
    $idx = 0;
    $count = count($this->parent->children);
    while ($idx<$count && $this! ==$this->parent->children[$idx])
        ++$idx;
    if (--$idx<0) return null;
    return $this->parent->children[$idx];
}
// get dom node's inner html
function innertext () {
    if (isset($this->_[HDOM_INFO_INNER])) return $this->_[HDOM_INFO_INNER];
    if (isset($this->_[HDOM_INFO_TEXT])) return $this->dom->restore_noise($this->_[HDOM_INFO_TEXT]);
    $ret = "";
    foreach($this->nodes as $n)
        $ret.= $n->outertext ();
    return $ret;
}
// get dom node's outer text (with tag)
function outertext () {
    if ($this->tag==='root') return $this->innertext ();
    // trigger callback
    if ($this->dom->callback! ==null)
        call_user_func_array($this->dom->callback, array($this));
    if (isset($this->_[HDOM_INFO_OUTER])) return $this->_[HDOM_INFO_OUTER];
    if (isset($this->_[HDOM_INFO_TEXT])) return $this->dom->restore_noise($this->_[HDOM_INFO_TEXT]);
}

```

```

// render begin tag
$ret = $this->dom->nodes[$this->_[HDOM_INFO_BEGIN]]->makeup ();
// render inner text
if (isset($this->_[HDOM_INFO_INNER]))
    $ret.= $this->_[HDOM_INFO_INNER];
else {
    foreach($this->nodes as $n)
        $ret.= $n->outertext ();
}
// render end tag
if(isset($this->_[HDOM_INFO_END]) && $this->_[HDOM_INFO_END]! =0)
    $ret.= '</'. $this->tag.'>';
return $ret;
}
// get dom node's plain text
function text () {
    if (isset($this->_[HDOM_INFO_INNER])) return $this->_[HDOM_INFO_INNER];
    switch ($this->nodetype) {
        case HDOM_TYPE_TEXT: return $this->dom->restore_noise($this->_[HDOM_INFO_TEXT]);
        case HDOM_TYPE_COMMENT: return "";
        case HDOM_TYPE_UNKNOWN: return "";
    }
    if (strcasecmp($this->tag, 'script') ===0) return "";
    if (strcasecmp($this->tag, 'style') ===0) return "";
    $ret = "";
    foreach($this->nodes as $n)
        $ret.= $n->text ();
    return $ret;
}
function xmltext () {
    $ret = $this->innertext ();
    $ret = str_ireplace ('<![CDATA [', '', $ret);
    $ret = str_replace (']>', '', $ret);
    return $ret;
}

```

```

}
// build node's text with tag
function makeup () {
    // text, comment, unknown
    if (isset($this->[_HDOM_INFO_TEXT])) return $this->dom->restore_noise($this->[_HDOM_INFO_TEXT]);
    $ret = '< ' . $this->tag;
    $i = -1;
    foreach($this->attr as $key=>$val) {
        ++$i;
        // skip removed attribute
        if ($val===null || $val===false)
            continue;
        $ret.= $this->[_HDOM_INFO_SPACE] [$i] [0];
        //no value attr: nowrap, checked selected...
        if ($val===true)
            $ret.= $key;
        else {
            switch($this->[_HDOM_INFO_QUOTE] [$i]) {
                case HDOM_QUOTE_DOUBLE: $quote = '"'; break;
                case HDOM_QUOTE_SINGLE: $quote = "'"; break;
                default: $quote = " ";
            }
            $ret.= $key. $this->[_HDOM_INFO_SPACE] [$i] [1].' '. $this->[_HDOM_INFO_SPACE] [$i] [2]. $quote. $val.$quote;
        }
    }
    $ret = $this->dom->restore_noise($ret);
    return $ret. $this->[_HDOM_INFO_ENDSPACE]. '>';
}

// find elements by css selector
function find ($selector, $idx=null) {
    $selectors = $this->parse_selector($selector);
    if (($count=count($selectors)) ===0) return array ();
    $found_keys = array ();

```

```

// find each selector
for ($c=0; $c<$count; ++$c) {
    if (($levle=count ($selectors [0])) ===0) return array ();
    if (! isset($this->_[HDOM_INFO_BEGIN])) return array ();
    $head = array($this->_[HDOM_INFO_BEGIN] =>1);
    // handle descendant selectors, no recursive!
    for ($l=0; $l<$levle; ++$l) {
        $ret = array ();
        foreach ($head as $k=>$v) {
            $n = ($k===-1)? $this->dom->root: $this->dom->nodes[$k];
            $n->seek($selectors[$c] [$l], $ret);
        }
        $head = $ret;
    }
    foreach ($head as $k=>$v) {
        if (! isset($found_keys[$k]))
            $found_keys[$k] = 1;
    }
}
// sort keys
ksort($found_keys);
$found = array ();
foreach ($found_keys as $k=>$v)
    $found [] = $this->dom->nodes[$k];
// return nth-element or array
if (is_null($idx)) return $found;
    else if ($idx<0) $idx = count($found) + $idx;
return (isset($found[$idx]))? $found[$idx]: null;
}
// seek for given conditions
protected function seek ($selector, &$ret) {
    list ($tag, $key, $val, $exp, $no_key) = $selector;
    // xpath index
    if ($tag && $key && is_numeric($key)) {
        $count = 0;

```

```

foreach ($this->children as $c) {
    if ($tag==='*' || $tag=== $c->tag) {
        if (++$count==$key) {
            $ret[$c->_[HDOM_INFO_BEGIN]] = 1;
            return;
        }
    }
}
return;
}
$end = (! empty($this->_[HDOM_INFO_END]))? $this->_[HDOM_INFO_END]: 0;
if ($end==0) {
    $parent = $this->parent;
    while (! isset($parent->_[HDOM_INFO_END]) && $parent! ==null) {
        $end -= 1;
        $parent = $parent->parent;
    }
    $end += $parent->_[HDOM_INFO_END];
}
for($i=$this->_[HDOM_INFO_BEGIN] +1; $i<$end; ++$i) {
    $node = $this->dom->nodes[$i];
    $pass = true;
    if ($tag==='*' &&! $key) {
        if (in_array ($node, $this->children, true))
            $ret[$i] = 1;
        continue;
    }
    // compare tag
    if ($tag && $tag! =$node->tag && $tag! =='*') {$pass=false;}
    // compare key
    if ($pass && $key) {
        if ($no_key) {
            if (isset($node->attr[$key])) $pass=false;
        }
        else if (! isset($node->attr[$key])) $pass=false;
    }
}

```

```

    }
    // compare value
    if ($pass && $key && $val && $val! == '*') {
        $check = $this->match ($exp, $val, $node->attr[$key]);
        // handle multiple class
        if (! $check && strcasecmp ($key, 'class') === 0) {
            foreach (explode (' ', $node->attr[$key]) as $k) {
                $check = $this->match ($exp, $val, $k);
                if ($check) break;
            }
        }
        if (! $check) $pass = false;
    }
    if ($pass) $ret[$i] = 1;
    unset($node);
}
}
protected function match ($exp, $pattern, $value) {
    switch ($exp) {
        case '=':
            return ($value=== $pattern);
        case '! =':
            return ($value! == $pattern);
        case '^=':
            return preg_match ("/^?". preg_quote($pattern, '/') ."/", $value);
        case '$=':
            return preg_match ("/?". preg_quote($pattern, '/') ."$"/, $value);
        case '*=':
            if ($pattern [0] == '/')
                return preg_match ($pattern, $value);
            return preg_match ("/?". $pattern ."/i", $value);
    }
    return false;
}
protected function parse_selector($selector_string) {

```

```

// pattern of CSS selectors, modified from mootools
$pattern = "/ ([\w-: \*] *) (? \#([\w-] +) \. ([\w-] +))? (? \@? (![\w-] +) (? :([! *^$]? =)
[\"']?(.*?)[\"']?\?)?([\v, ]+)/is";
preg_match_all ($pattern, trim($selector_string).' ', $matches, PREG_SET_ORDER);
$selectors = array ();
$result = array ();
//print_r($matches);
foreach ($matches as $m) {
    $m [0] = trim ($m [0]);
    if ($m [0] ===" || $m[0]==='" || $m[0]==='/') continue;
    // for borwser grnreated xpath
    if ($m [1] ==='tbody') continue;
    list ($tag, $key, $val, $exp, $no_key) = array ($m [1], null, null, '=', false);
    if (! empty ($m [2])) {$key='id'; $val=$m [2];}
    if (! empty ($m [3])) {$key='class'; $val=$m [3];}
    if (! empty ($m [4])) {$key=$m [4];}
    if (! empty ($m [5])) {$exp=$m [5];}
    if (! empty ($m [6])) {$val=$m [6];}
    // convert to lowercase
    if ($this->dom->lowercase) {$tag=strtolower($tag); $key=strtolower($key);}
    //elements that do NOT have the specified attribute
    if (isset ($key [0]) && $key [0] ==='!') {$key=substr ($key, 1); $no_key=true;}
    $result [] = array ($tag, $key, $val, $exp, $no_key);
    if (trim ($m [7]) ===',') {
        $selectors [] = $result;
        $result = array ();
    }
}
if (count($result)>0)
    $selectors [] = $result;
return $selectors;
}
function __get($name) {
    if (isset($this->attr[$name])) return $this->attr[$name];
    switch($name) {

```

```

        case 'outertext': return $this->outertext ();
        case 'innertext': return $this->innertext ();
        case 'plaintext': return $this->text ();
        case 'xmltext': return $this->xmltext ();
        default: return array_key_exists ($name, $this->attr);
    }
}
function __set ($name, $value) {
    switch($name) {
        case 'outertext': return $this->_[HDOM_INFO_OUTER] = $value;
        case 'innertext':
            if (isset($this->_[HDOM_INFO_TEXT])) return $this->_[HDOM_INFO_TEXT] =
$value;
            return $this->_[HDOM_INFO_INNER] = $value;
        }
        if (! isset($this->attr[$name])) {
            $this->_[HDOM_INFO_SPACE] [] = array (' ', ", ");
            $this->_[HDOM_INFO_QUOTE] [] = HDOM_QUOTE_DOUBLE;
        }
        $this->attr[$name] = $value;
    }
}
function __isset($name) {
    switch($name) {
        case 'outertext': return true;
        case 'innertext': return true;
        case 'plaintext': return true;
    }
    //no value attr: nowrap, checked selected...
    return (array_key_exists ($name, $this->attr)? true: isset($this->attr[$name]));
}
function __unset($name) {
    if (isset($this->attr[$name]))
        unset($this->attr[$name]);
}
// camel naming conventions

```

```

function getAllAttributes () {return $this->attr;}
function getAttribute($name) {return $this->__get($name);}
function setAttribute ($name, $value) {$this->__set ($name, $value);}
function hasAttribute($name) {return $this->__isset($name);}
function removeAttribute($name) {$this->__set ($name, null);}
function getElementById($id) {return $this->find ("#$id", 0);}
function getElementsById ($id, $idx=null) {return $this->find ("#$id", $idx);}
function getElementByTagName($name) {return $this->find ($name, 0);}
function getElementsByTagName ($name, $idx=null) {return $this->find ($name, $idx);}
function parentNode () {return $this->parent ();}
function childNodes($idx=-1) {return $this->children($idx);}
function firstChild () {return $this->first_child ();}
function lastChild () {return $this->last_child ();}
function nextSibling () {return $this->next_sibling ();}
function previousSibling () {return $this->prev_sibling ();}
}
// simple html dom parser
// -----
class simple_html_dom {
    public $root = null;
    public $nodes = array ();
    public $callback = null;
    public $lowercase = false;
    protected $pos;
    protected $doc;
    protected $char;
    protected $size;
    protected $cursor;
    protected $parent;
    protected $noise = array ();
    protected $token_blank = "\t\r\n";
    protected $token_equal = ' =/>';
    protected $token_slash = " />\r\n\t";
    protected $token_attr = '>';
    // use isset instead of in_array, performance boost about 30% ...

```

```

protected $self_closing_tags = array('img'=>1, 'br'=>1, 'input'=>1, 'meta'=>1, 'link'=>1,
'hr'=>1, 'base'=>1, 'embed'=>1, 'spacer'=>1);
protected $block_tags = array('root'=>1, 'body'=>1, 'form'=>1, 'div'=>1, 'span'=>1,
'table'=>1);
protected $optional_closing_tags = array (
'tr'=>array('tr'=>1, 'td'=>1, 'th'=>1),
'th'=>array('th'=>1),
'td'=>array('td'=>1),
'li'=>array('li'=>1),
'dt'=>array('dt'=>1, 'dd'=>1),
'dd'=>array('dd'=>1, 'dt'=>1),
'dl'=>array('dd'=>1, 'dt'=>1),
'p'=>array('p'=>1),
'nobr'=>array('nobr'=>1),
);
function __construct($str=null) {
if ($str) {
if (preg_match ("/^http: \\/\\/i", $str) || is_file($str))
$this->load_file($str);
else
$this->load($str);
}
}
function __destruct () {
$this->clear ();
}
// load html from string
function load ($str, $lowercase=true) {
// prepare
$this->prepare ($str, $lowercase);
// strip out comments
$this->remove_noise ("<! --(. *) -->'is");
// strip out cdata
$this->remove_noise ("<! \ [CDATA\ [(. *) \ ]>'is", true);
// strip out <style> tags

```

```

$this->remove_noise("'<\s*style[^>]*[^\>](. *?)<\s*/\s*style\s*>'is");
$this->remove_noise("'<\s*style\s*>(.*?)<\s*/\s*style\s*>'is");
// strip out <script> tags
$this->remove_noise ("'\<\s*script [^>] *[^>]> (. *?) <\s*/\s*script\s*>'is");
$this->remove_noise("'<\s*script\s*> (. *?) <\s*/\s*script\s*>'is");
// strip out preformatted tags
$this->remove_noise ("'\<\s*(? code) [^>] *> (. *?) <\s*/\s*(? code) \s*>'is");
// strip out server side scripts
$this->remove_noise ("'(<\/?) (. *?) (\/>)'s", true);
// strip smarty scripts
$this->remove_noise ("'(\{\w) (. *?) (\})'s", true);
// parsing
while ($this->parse ());
// end
$this->root->[_HDOM_INFO_END] = $this->cursor;
}
// load html from file
function load_file () {
    $args = func_get_args ();
    $this->load (call_user_func_array ('file_get_contents', $args), true);
}
// set callback function
function set_callback($function_name) {
    $this->callback = $function_name;
}
// remove callback function
function remove_callback () {
    $this->callback = null;
}
// save dom as string
function save($filepath=") {
    $ret = $this->root->innertext ();
    if ($filepath! ==") file_put_contents ($filepath, $ret);
    return $ret;
}

```

```

// find dom node by css selector
function find ($selector, $idx=null) {
    return $this->root->find ($selector, $idx);
}
// clean up memory due to php5 circular references memory leak...
function clear () {
    foreach($this->nodes as $n) {$n->clear (); $n = null;}
    if (isset($this->parent)) {$this->parent->clear (); unset($this->parent);}
    if (isset($this->root)) {$this->root->clear (); unset($this->root);}
    unset($this->doc);
    unset($this->noise);
}

function dump($show_attr=true) {
    $this->root->dump($show_attr);
}
// prepare HTML data and init everything
protected function prepare ($str, $lowercase=true) {
    $this->clear ();
    $this->doc = $str;
    $this->pos = 0;
    $this->cursor = 1;
    $this->noise = array ();
    $this->nodes = array ();
    $this->lowercase = $lowercase;
    $this->root = new simple_html_dom_node($this);
    $this->root->tag = 'root';
    $this->root->_[HDOM_INFO_BEGIN] = -1;
    $this->root->nodetype = HDOM_TYPE_ROOT;
    $this->parent = $this->root;
    // set the length of content
    $this->size = strlen($str);
    if ($this->size>0) $this->char = $this->doc [0];
}
// parse html content

```

```

protected function parse () {
    if (($s = $this->copy_until_char ('<')) === "")
        return $this->read_tag ();
    // text
    $node = new simple_html_dom_node($this);
    ++$this->cursor;
    $node->[_HDOM_INFO_TEXT] = $s;
    $this->link_nodes ($node, false);
    return true;
}
// read tag info
protected function read_tag () {
    if ($this->char! == '<') {
        $this->root->[_HDOM_INFO_END] = $this->cursor;
        return false;
    }
    $begin_tag_pos = $this->pos;
    $this->char = (++$this->pos < $this->size)? $this->doc[$this->pos]: null; // next
    // end tag
    if ($this->char === '/') {
        $this->char = (++$this->pos < $this->size)? $this->doc[$this->pos]: null; // next
        $this->skip($this->token_blank_t);
        $tag = $this->copy_until_char ('>');
        // skip attributes in end tag
        if (($pos = strpos ($tag, ' ')) !== false)
            $tag = substr ($tag, 0, $pos);
        $parent_lower = strtolower($this->parent->tag);
        $tag_lower = strtolower($tag);
        if ($parent_lower! == $tag_lower) {
            if (isset($this->optional_closing_tags[$parent_lower]) && isset($this->block_tags[$tag_lower])) {
                $this->parent->[_HDOM_INFO_END] = 0;
                $org_parent = $this->parent;
                while (($this->parent->parent) && strtolower($this->parent->tag)! == $tag_lower)

```

```

        $this->parent = $this->parent->parent;
    if (strtolower($this->parent->tag) !== $tag_lower) {
        $this->parent = $org_parent; // restore original parent
        if ($this->parent->parent) $this->parent = $this->parent->parent;
        $this->parent->_[HDOM_INFO_END] = $this->cursor;
        return $this->as_text_node($tag);
    }
}
else if (($this->parent->parent) && isset($this->block_tags[$tag_lower])) {
    $this->parent->_[HDOM_INFO_END] = 0;
    $org_parent = $this->parent;
    while (($this->parent->parent) && strtolower($this->parent->parent->tag) !== $tag_lower)
        $this->parent = $this->parent->parent;
    if (strtolower($this->parent->tag) !== $tag_lower) {
        $this->parent = $org_parent; // restore original parent
        $this->parent->_[HDOM_INFO_END] = $this->cursor;
        return $this->as_text_node($tag);
    }
}
else if (($this->parent->parent) && strtolower($this->parent->parent->tag) === $tag_lower) {
    $this->parent->_[HDOM_INFO_END] = 0;
    $this->parent = $this->parent->parent;
}
else
    return $this->as_text_node($tag);
}
$this->parent->_[HDOM_INFO_END] = $this->cursor;
if ($this->parent->parent) $this->parent = $this->parent->parent;
$this->char = (++$this->pos < $this->size)? $this->doc[$this->pos] : null; // next
return true;
}
}
$node = new simple_html_dom_node($this);
$node->_[HDOM_INFO_BEGIN] = $this->cursor;

```

```

++$this->cursor;
$tag = $this->copy_until($this->token_slash);
// doctype, cdata & comments...
if (isset ($tag [0]) && $tag [0] === '!') {
    $node->_[HDOM_INFO_TEXT] = '< ' . $tag . $this->copy_until_char ('>');
    if (isset ($tag [2]) && $tag [1] === '-' && $tag [2] === '-') {
        $node->nodetype = HDOM_TYPE_COMMENT;
        $node->tag = 'comment';
    } else {
        $node->nodetype = HDOM_TYPE_UNKNOWN;
        $node->tag = 'unknown';
    }
    if ($this->char====>) $node->_[HDOM_INFO_TEXT].='>';
    $this->link_nodes ($node, true);
    $this->char = (++$this->pos<$this->size) ? $this->doc[$this->pos] : null; // next
    return true;
}
// text
if ($pos=strpos ($tag, '<')! ==false) {
    $tag = '< ' . substr ($tag, 0, -1);
    $node->_[HDOM_INFO_TEXT] = $tag;
    $this->link_nodes ($node, false);
    $this->char = $this->doc[--$this->pos]; // prev
    return true;
}
if (! preg_match("/^\[w-:]+$/", $tag)) {
    $node->_[HDOM_INFO_TEXT] = '<' . $tag . $this->copy_until ('<>');
    if ($this->char===<) {
        $this->link_nodes ($node, false);
        return true;
    }
    if ($this->char====>) $node->_[HDOM_INFO_TEXT].='>';
    $this->link_nodes ($node, false);
    $this->char = (++$this->pos<$this->size)? $this->doc[$this->pos]: null; // next
    return true;
}

```

```

}
// begin tag
$node->nodetype = HDOM_TYPE_ELEMENT;
$tag_lower = strtolower($tag);
$node->tag = ($this->lowercase)? $tag_lower: $tag;
// handle optional closing tags
if (isset($this->optional_closing_tags[$tag_lower])) {
    while (isset($this->optional_closing_tags[$tag_lower] [strtolower($this->parent-
>tag)))) {
        $this->parent->_[HDOM_INFO_END] = 0;
        $this->parent = $this->parent->parent;
    }
    $node->parent = $this->parent;
}
$guard = 0; // prevent infinity loop
$space = array($this->copy_skip($this->token_blank), ", ");
// attributes
do {
    if ($this->char! ==null && $space [0] ==") break;
    $name = $this->copy_until($this->token_equal);
    if($guard=== $this->pos) {
        $this->char = (++$this->pos<$this->size)? $this->doc[$this->pos]: null; // next
        continue;
    }
    $guard = $this->pos;
    // handle endless '<'
    if($this->pos>=$this->size-1 && $this->char! ==>') {
        $node->nodetype = HDOM_TYPE_TEXT;
        $node->_[HDOM_INFO_END] = 0;
        $node->_[HDOM_INFO_TEXT] = '<'. $tag . $space [0] . $name;
        $node->tag = 'text';
        $this->link_nodes ($node, false);
        return true;
    }
    // handle mismatch '<'

```

```

if($this->doc[$this->pos-1] == '<') {
    $node->nodetype = HDOM_TYPE_TEXT;
    $node->tag = 'text';
    $node->attr = array ();
    $node->_[HDOM_INFO_END] = 0;
    $node->_[HDOM_INFO_TEXT] = substr($this->doc, $begin_tag_pos, $this->pos-
$begin_tag_pos-1);
    $this->pos -= 2;
    $this->char = (++$this->pos < $this->size)? $this->doc[$this->pos]: null; // next
    $this->link_nodes ($node, false);
    return true;
}
if ($name!=='/' && $name!=='') {
    $space [1] = $this->copy_skip($this->token_blank);
    $name = $this->restore_noise($name);
    if ($this->lowercase) $name = strtolower($name);
    if ($this->char==='=') {
        $this->char = (++$this->pos < $this->size)? $this->doc[$this->pos]: null; // next
        $this->parse_attr ($node, $name, $space);
    }
    else {
        //no value attr: nowrap, checked selected...
        $node->_[HDOM_INFO_QUOTE] [] = HDOM_QUOTE_NO;
        $node->attr[$name] = true;
        if ($this->char!=='>') $this->char = $this->doc[--$this->pos]; // prev
    }
    $node->_[HDOM_INFO_SPACE] [] = $space;
    $space = array($this->copy_skip($this->token_blank), " ", "");
}
else
    break;
} while($this->char!=='>' && $this->char!=='/');
$this->link_nodes ($node, true);
$node->_[HDOM_INFO_ENDSPACE] = $space [0];
// check self closing

```

```

if ($this->copy_until_char_escape ('> ' === '/') {
    $node->_[HDOM_INFO_ENDSPACE]. = '/';
    $node->_[HDOM_INFO_END] = 0;
}
else {
    // reset parent
    if (! isset($this->self_closing_tags[strtolower($node->tag)]) $this->parent = $node;
}
$this->char = (++$this->pos<$this->size)? $this->doc[$this->pos]: null; // next
return true;
}
// parse attributes
protected function parse_attr ($node, $name, &$space) {
    $space [2] = $this->copy_skip($this->token_blank);
    switch($this->char) {
        case '"':
            $node->_[HDOM_INFO_QUOTE] [] = HDOM_QUOTE_DOUBLE;
            $this->char = (++$this->pos<$this->size)? $this->doc[$this->pos]: null; // next
            $node->attr[$name] = $this->restore_noise($this->copy_until_char_escape ('"));
            $this->char = (++$this->pos<$this->size)? $this->doc[$this->pos]: null; // next
            break;
        case '\':
            $node->_[HDOM_INFO_QUOTE] [] = HDOM_QUOTE_SINGLE;
            $this->char = (++$this->pos<$this->size)? $this->doc[$this->pos]: null; // next
            $node->attr[$name] = $this->restore_noise($this->copy_until_char_escape ('\'));
            $this->char = (++$this->pos<$this->size)? $this->doc[$this->pos]: null; // next
            break;
        default:
            $node->_[HDOM_INFO_QUOTE] [] = HDOM_QUOTE_NO;
            $node->attr[$name] = $this->restore_noise($this->copy_until($this->token_attr));
    }
}
// link node's parent
protected function link_nodes (&$node, $is_child) {
    $node->parent = $this->parent;
}

```

```

    $this->parent->nodes [] = $node;
    if ($is_child)
        $this->parent->children [] = $node;
}
// as a text node
protected function as_text_node($tag) {
    $node = new simple_html_dom_node($this);
    ++$this->cursor;
    $node->[_HDOM_INFO_TEXT] = '</'. $tag. '>';
    $this->link_nodes ($node, false);
    $this->char = (++$this->pos<$this->size)? $this->doc[$this->pos]: null; // next
    return true;
}
protected function skip($chars) {
    $this->pos += strspn($this->doc, $chars, $this->pos);
    $this->char = ($this->pos<$this->size) ? $this->doc[$this->pos] : null; // next
}
protected function copy_skip($chars) {
    $pos = $this->pos;
    $len = strspn($this->doc, $chars, $pos);
    $this->pos += $len;
    $this->char = ($this->pos<$this->size) ? $this->doc[$this->pos] : null; // next
    if ($len===0) return "";
    return substr($this->doc, $pos, $len);
}
protected function copy_until($chars) {
    $pos = $this->pos;
    $len = strcspn($this->doc, $chars, $pos);
    $this->pos += $len;
    $this->char = ($this->pos<$this->size) ? $this->doc[$this->pos] : null; // next
    return substr($this->doc, $pos, $len);
}
protected function copy_until_char($char) {
    if ($this->char===null) return "";
    if (($pos = strpos($this->doc, $char, $this->pos))===false) {

```

```

$ret = substr($this->doc, $this->pos, $this->size-$this->pos);
$this->char = null;
$this->pos = $this->size;
return $ret;
}
if ($pos=== $this->pos) return "";
$pos_old = $this->pos;
$this->char = $this->doc[$pos];
$this->pos = $pos;
return substr($this->doc, $pos_old, $pos-$pos_old);
}
protected function copy_until_char_escape($char) {
if ($this->char===null) return "";
$start = $this->pos;
while (1) {
if (($pos = strpos($this->doc, $char, $start)) ===false) {
$ret = substr($this->doc, $this->pos, $this->size-$this->pos);
$this->char = null;
$this->pos = $this->size;
return $ret;
}
if ($pos=== $this->pos) return "";
if ($this->doc[$pos-1] ==="\") {
$start = $pos+1;
continue;
}
}
$pos_old = $this->pos;
$this->char = $this->doc[$pos];
$this->pos = $pos;
return substr($this->doc, $pos_old, $pos-$pos_old);
}
}
// remove noise from html content
protected function remove_noise ($pattern, $remove_tag=false) {

```

```

    $count = preg_match_all ($pattern, $this->doc, $matches,
PREG_SET_ORDER|PREG_OFFSET_CAPTURE);
    for ($i=$count-1; $i>-1; --$i) {
        $key = '___noise___'. sprintf ('% 3d', count($this->noise) +100);
        $idx = ($remove_tag)? 0: 1;
        $this->noise[$key] = $matches[$i] [$idx] [0];
        $this->doc = substr_replace($this->doc, $key, $matches[$i] [$idx] [1],
strlen($matches[$i] [$idx] [0]));
    }
    // reset the length of content
    $this->size = strlen($this->doc);
    if ($this->size>0) $this->char = $this->doc [0];
}
// restore noise to html content
function restore_noise($text) {
    while (($pos=strpos ($text, '___noise___'))! ==false) {
        $key = '___noise___'. $text[$pos+11]. $text[$pos+12]. $text[$pos+13];
        if (isset($this->noise[$key]))
            $text = substr ($text, 0, $pos). $this->noise[$key]. substr ($text, $pos+14);
    }
    return $text;
}
function __toString () {
    return $this->root->innertext ();
}
function __get($name) {
    switch($name) {
        case 'outertext': return $this->root->innertext ();
        case 'innertext': return $this->root->innertext ();
        case 'plaintext': return $this->root->text ();
    }
}
}
// camel naming conventions
function childNodes($idx=-1) {return $this->root->childNodes($idx);}
function firstChild () {return $this->root->first_child ();}

```

```

function lastChild () {return $this->root->last_child ();}
function getElementById($id) {return $this->find ("#$id", 0);}
function getElementsById ($id, $idx=null) {return $this->find ("#$id", $idx);}
function getElementByTagName($name) {return $this->find ($name, 0);}
function getElementsByTagName ($name, $idx=-1) {return $this->find ($name, $idx);}
function loadFile () {$args = func_get_args (); $this->load (call_user_func_array
('file_get_contents', $args), true);}
}
?>

```

Extrayendo datos con el MySQL Administrador

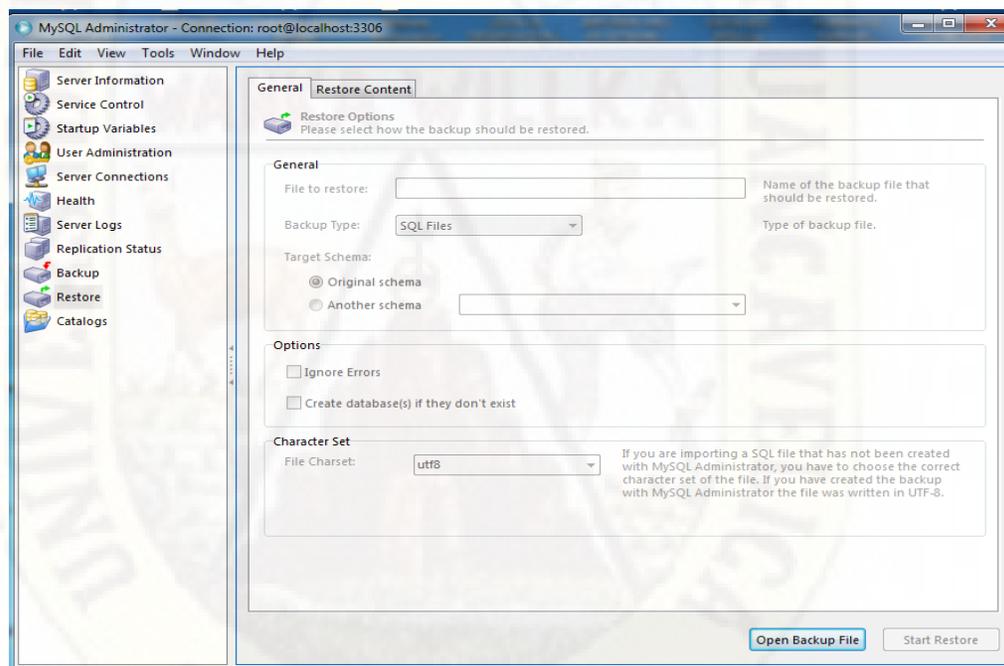


Figura 33. Extracción de datos

Fuente: Elaboración propia

Exportando datos con el Navicat 8 for MySQL

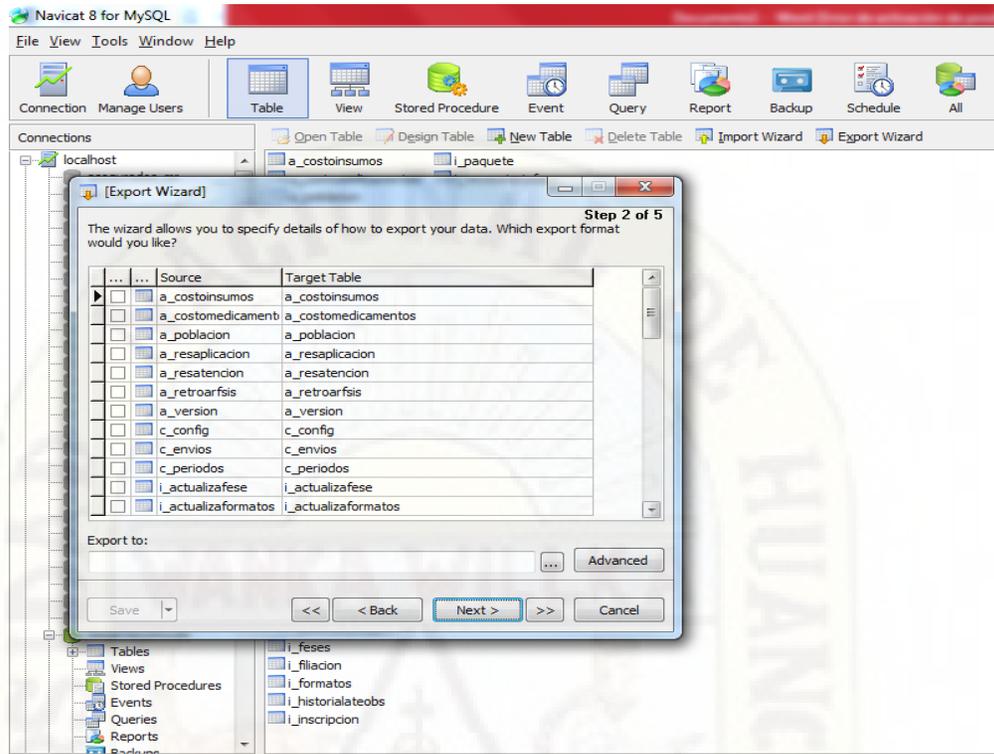


Figura 34. Exportación de datos

Fuente: Elaboración propia

Diseño de la Base de Datos

Tabla de base de datos_Ate (atenciones) del sistema Web Scraping

| Nombre de columna | Tipo de datos | Permitir valores NULL |
|-----------------------|---------------|-------------------------------------|
| P_Digi | nvarchar(255) | <input checked="" type="checkbox"/> |
| ate_IdAtencion | int | <input checked="" type="checkbox"/> |
| ate_IdFormato | int | <input checked="" type="checkbox"/> |
| ate_IdDisa | nvarchar(9) | <input checked="" type="checkbox"/> |
| ate_Lote | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Numero | nvarchar(24) | <input checked="" type="checkbox"/> |
| ate_EsReconsideracion | nvarchar(3) | <input checked="" type="checkbox"/> |
| ate_IdDisaReconsidera | nvarchar(9) | <input checked="" type="checkbox"/> |
| ate_LoteReconsidera | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_NroReconsidera | nvarchar(24) | <input checked="" type="checkbox"/> |
| ate_IdEESS | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_IdComponente | nvarchar(3) | <input checked="" type="checkbox"/> |
| ate_IdTipoFormato | int | <input checked="" type="checkbox"/> |
| ate_IdSitFormato | int | <input checked="" type="checkbox"/> |
| ate_IdDisaFormato | nvarchar(9) | <input checked="" type="checkbox"/> |
| ate_LoteFormato | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_NroFormato | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_IdTipoDocumento | nvarchar(3) | <input checked="" type="checkbox"/> |
| ate_NroDocumento | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_IdInstitucion | nvarchar(9) | <input checked="" type="checkbox"/> |
| ate_NroInstitucion | nvarchar(90) | <input checked="" type="checkbox"/> |
| ate_idIIIE | nvarchar(24) | <input checked="" type="checkbox"/> |

Figura 35. Tabla atenciones de la base de datos del modelo de sistema Web Scraping

Fuente: Elaboracion propia

| | | |
|----------------------|---------------|-------------------------------------|
| ate_idnivelIIIE | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_idgradoIIIE | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_seccionIIIE | nvarchar(60) | <input checked="" type="checkbox"/> |
| ate_idturnoIIIE | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_ApePaterno | nvarchar(120) | <input checked="" type="checkbox"/> |
| ate_ApeMaterno | nvarchar(120) | <input checked="" type="checkbox"/> |
| ate_PriNombre | nvarchar(105) | <input checked="" type="checkbox"/> |
| ate_OtrNombre | nvarchar(135) | <input checked="" type="checkbox"/> |
| ate_FecNac | datetime | <input checked="" type="checkbox"/> |
| ate_IdSexo | nvarchar(3) | <input checked="" type="checkbox"/> |
| ate_IdTipoAtencion | nvarchar(3) | <input checked="" type="checkbox"/> |
| ate_IdCondicion | nvarchar(3) | <input checked="" type="checkbox"/> |
| ate_IdModalidad | nvarchar(3) | <input checked="" type="checkbox"/> |
| ate_NroAutorizacion | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_Monto | float | <input checked="" type="checkbox"/> |
| ate_FecAte | datetime | <input checked="" type="checkbox"/> |
| ate_HoraAte | datetime | <input checked="" type="checkbox"/> |
| ate_IdLugar | nvarchar(3) | <input checked="" type="checkbox"/> |
| ate_HisCli | nvarchar(60) | <input checked="" type="checkbox"/> |
| ate_FecParto | datetime | <input checked="" type="checkbox"/> |
| ate_IdOrigenPersonal | nvarchar(3) | <input checked="" type="checkbox"/> |

Figura 36. Tabla atenciones de la base de datos del modelo de sistema Web Scraping

Fuente: Elaboracion propia

| | | |
|--------------------------|---------------|-------------------------------------|
| ate_IdServicio | nvarchar(9) | <input checked="" type="checkbox"/> |
| ate_IdEESSRefirio | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_NroHojaRefirio | nvarchar(60) | <input checked="" type="checkbox"/> |
| ate_IdDestinoAsegurado | nvarchar(3) | <input checked="" type="checkbox"/> |
| ate_IdEESSContraRefiere | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_NroHojaContraRefiere | nvarchar(60) | <input checked="" type="checkbox"/> |
| ate_FecIng | datetime | <input checked="" type="checkbox"/> |
| ate_FecAlta | datetime | <input checked="" type="checkbox"/> |
| ate_FecFallecimiento | datetime | <input checked="" type="checkbox"/> |
| ate_IdResAtencion | nvarchar(27) | <input checked="" type="checkbox"/> |
| ate_Profesional | nvarchar(255) | <input checked="" type="checkbox"/> |
| ate_IdTipoPersonalSalud | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Colegiatura | nvarchar(150) | <input checked="" type="checkbox"/> |
| ate_IdEspecialidad | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_NroEspecialidad | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_Observacion | nvarchar(255) | <input checked="" type="checkbox"/> |
| ate_Edad | int | <input checked="" type="checkbox"/> |
| ate_TipoEdad | int | <input checked="" type="checkbox"/> |
| ate_EdadReal | int | <input checked="" type="checkbox"/> |
| ate_Edades | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_IdGrupoEtereo | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Autogenerado | nvarchar(60) | <input checked="" type="checkbox"/> |

Figura 37. abla atenciones de la base de datos del modelo de sistema Web Scraping

Fuente: Elaboracion propia

| | | |
|-----------------------|--------------|-------------------------------------|
| ate_IdPPDD | int | <input checked="" type="checkbox"/> |
| ate_IdOdsis | nvarchar(9) | <input checked="" type="checkbox"/> |
| ate_IdUsuarioTrans | nvarchar(24) | <input checked="" type="checkbox"/> |
| ate_FecTrans | datetime | <input checked="" type="checkbox"/> |
| ate_IdGrupoRiesgo | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_CostoServ | float | <input checked="" type="checkbox"/> |
| ate_CostoMedi | float | <input checked="" type="checkbox"/> |
| ate_CostoInsu | float | <input checked="" type="checkbox"/> |
| ate_CostoProc | float | <input checked="" type="checkbox"/> |
| ate_CostoTotal | float | <input checked="" type="checkbox"/> |
| ate_Periodo | nvarchar(12) | <input checked="" type="checkbox"/> |
| ate_Mes | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Envio | int | <input checked="" type="checkbox"/> |
| ate_IdSupervision | int | <input checked="" type="checkbox"/> |
| ate_IdEstado | nvarchar(3) | <input checked="" type="checkbox"/> |
| ate_Version | nvarchar(27) | <input checked="" type="checkbox"/> |
| ate_CorrelativoInsAus | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_IdFormatoSiasis | int | <input checked="" type="checkbox"/> |
| ate_TipoTablaSiasis | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_FecCrea | datetime | <input checked="" type="checkbox"/> |

Figura 38. abla atenciones de la base de datos del modelo de sistema Web Scraping

Fuente: Elaboracion propia

| | | |
|------------------------|---------------|-------------------------------------|
| ate_PerOrigen | nvarchar(18) | <input checked="" type="checkbox"/> |
| ate_codaplicativo | nvarchar(255) | <input checked="" type="checkbox"/> |
| ate_IdIAFAS | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_codsegIAFAS | nvarchar(60) | <input checked="" type="checkbox"/> |
| ate_NroAutorizaCartaG | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_montoCartaG | float | <input checked="" type="checkbox"/> |
| ate_UDRautorizaFUA | nvarchar(9) | <input checked="" type="checkbox"/> |
| ate_LoteautorizaFUA | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_secautorizaFUA | nvarchar(18) | <input checked="" type="checkbox"/> |
| ate_disaFUAVinculado | nvarchar(9) | <input checked="" type="checkbox"/> |
| ate_loteFUAVinculado | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_secFUAVinculado | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_idsepelio | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_IdUPS | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_IdEtnia | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_idEESFlexible | nvarchar(30) | <input checked="" type="checkbox"/> |
| ate_FecCorteAdm | datetime | <input checked="" type="checkbox"/> |
| ATE_ESESGRESADOPERS... | nvarchar(3) | <input checked="" type="checkbox"/> |
| ATE_PLAN | nvarchar(6) | <input checked="" type="checkbox"/> |
| ATE_GRUPOPOBLACION... | nvarchar(6) | <input checked="" type="checkbox"/> |

Propiedades de columna

Figura 39. Tabla atenciones de la base de datos del modelo de sistema Web Scraping

Fuente: Elaboracion propia

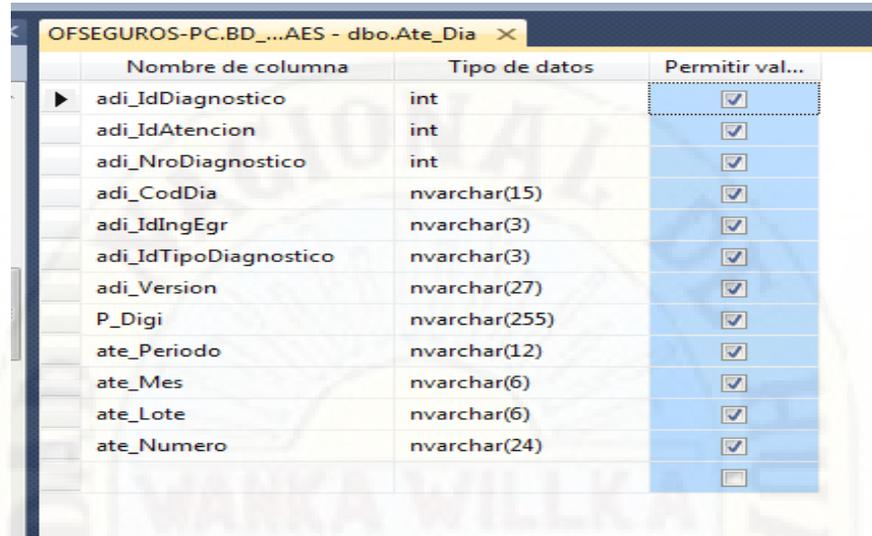
Tabla de base de datos_Insumos (insumos) del sistema Web Scraping

| Nombre de columna | Tipo de datos | Permitir val... |
|---------------------|---------------|-------------------------------------|
| ains_IdInsumo | int | <input checked="" type="checkbox"/> |
| ains_IdDiagnostico | int | <input checked="" type="checkbox"/> |
| ains_CodInsumo | nvarchar(15) | <input checked="" type="checkbox"/> |
| ains_CantPrescrita | int | <input checked="" type="checkbox"/> |
| ains_CantEntregada | int | <input checked="" type="checkbox"/> |
| ains_PrecioUnitario | float | <input checked="" type="checkbox"/> |
| ains_NroDiagnostico | int | <input checked="" type="checkbox"/> |
| ains_Version | nvarchar(27) | <input checked="" type="checkbox"/> |
| P_Digi | nvarchar(255) | <input checked="" type="checkbox"/> |
| ate_Periodo | nvarchar(12) | <input checked="" type="checkbox"/> |
| ate_Mes | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Lote | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Numero | nvarchar(24) | <input checked="" type="checkbox"/> |
| | | <input type="checkbox"/> |

Figura 40. Tabla insumos de la base de datos del modelo de sistema Web Scraping

Fuente: Elaboracion propia

Tabla de base de datos Ate_Dia (diagnosticos) del sistema Web Scraping



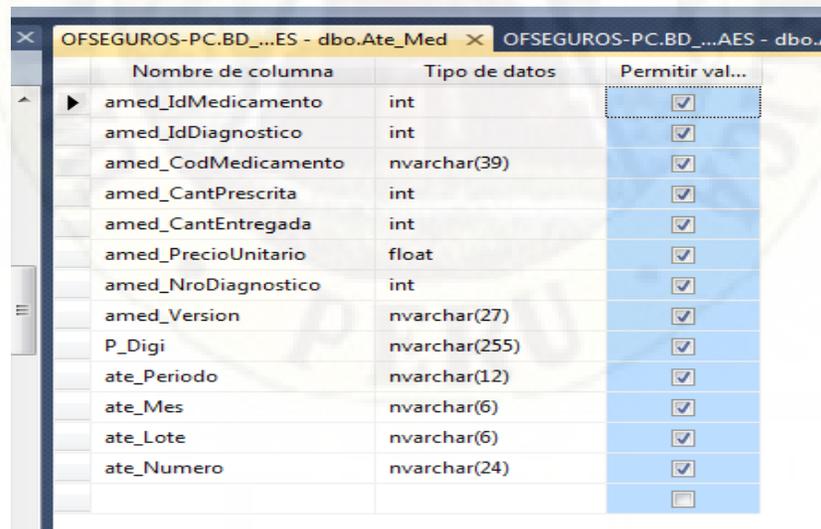
The screenshot shows a table named 'Ate_Dia' in the 'dbo' schema of the 'OFSEGUROS-PC.BD_...AES' database. The table has 13 columns with the following data types and 'Permitir valores NULL' status:

| Nombre de columna | Tipo de datos | Permitir val... |
|-----------------------|---------------|-------------------------------------|
| adi_IdDiagnostico | int | <input checked="" type="checkbox"/> |
| adi_IdAtencion | int | <input checked="" type="checkbox"/> |
| adi_NroDiagnostico | int | <input checked="" type="checkbox"/> |
| adi_CodDia | nvarchar(15) | <input checked="" type="checkbox"/> |
| adi_IdIngEgr | nvarchar(3) | <input checked="" type="checkbox"/> |
| adi_IdTipoDiagnostico | nvarchar(3) | <input checked="" type="checkbox"/> |
| adi_Version | nvarchar(27) | <input checked="" type="checkbox"/> |
| P_Digi | nvarchar(255) | <input checked="" type="checkbox"/> |
| ate_Periodo | nvarchar(12) | <input checked="" type="checkbox"/> |
| ate_Mes | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Lote | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Numero | nvarchar(24) | <input checked="" type="checkbox"/> |
| | | <input type="checkbox"/> |

Figura 41. Tabla diagnosticos de la base de datos del modelo de sistema Web Scraping

Fuente: Elaboracion propia

Tabla de base de datos Ate_Med (medicamentos) del sistema Web Scraping



The screenshot shows a table named 'Ate_Med' in the 'dbo' schema of the 'OFSEGUROS-PC.BD_...ES' database. The table has 14 columns with the following data types and 'Permitir valores NULL' status:

| Nombre de columna | Tipo de datos | Permitir val... |
|---------------------|---------------|-------------------------------------|
| amed_IdMedicamento | int | <input checked="" type="checkbox"/> |
| amed_IdDiagnostico | int | <input checked="" type="checkbox"/> |
| amed_CodMedicamento | nvarchar(39) | <input checked="" type="checkbox"/> |
| amed_CantPrescrita | int | <input checked="" type="checkbox"/> |
| amed_CantEntregada | int | <input checked="" type="checkbox"/> |
| amed_PrecioUnitario | float | <input checked="" type="checkbox"/> |
| amed_NroDiagnostico | int | <input checked="" type="checkbox"/> |
| amed_Version | nvarchar(27) | <input checked="" type="checkbox"/> |
| P_Digi | nvarchar(255) | <input checked="" type="checkbox"/> |
| ate_Periodo | nvarchar(12) | <input checked="" type="checkbox"/> |
| ate_Mes | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Lote | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Numero | nvarchar(24) | <input checked="" type="checkbox"/> |
| | | <input type="checkbox"/> |

Figura 42. Tabla medicamentos de la base de datos del modelo de sistema Web Scraping

Fuente: Elaboracion propia

Tabla base de datos Ate_Proced (procedimientos) del sistema Web Scraping

| Nombre de columna | Tipo de datos | Permitir val... |
|-----------------------|---------------|-------------------------------------|
| apro_IdProcedimiento | int | <input checked="" type="checkbox"/> |
| apro_IdDiagnostico | int | <input checked="" type="checkbox"/> |
| apro_CodProcedimiento | nvarchar(45) | <input checked="" type="checkbox"/> |
| apro_CantIndicado | int | <input checked="" type="checkbox"/> |
| apro_CantEjecutado | int | <input checked="" type="checkbox"/> |
| apro_PrecioUnitario | float | <input checked="" type="checkbox"/> |
| apro_NroDiagnostico | int | <input checked="" type="checkbox"/> |
| apro_Resultado | nvarchar(69) | <input checked="" type="checkbox"/> |
| apro_Version | nvarchar(27) | <input checked="" type="checkbox"/> |
| P_Digi | nvarchar(255) | <input checked="" type="checkbox"/> |
| ate_Periodo | nvarchar(12) | <input checked="" type="checkbox"/> |
| ate_Mes | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Lote | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Numero | nvarchar(24) | <input checked="" type="checkbox"/> |
| | | <input type="checkbox"/> |

Figura 43. Tabla procedimientos de la base de datos del modelo de sistema Web Scraping

Fuente: Elaboracion propia

Tabla base de datos Ate_SMI (servicio materno infantil) del sistema Web Scraping

| Nombre de columna | Tipo de datos | Permitir val... |
|-----------------------|---------------|-------------------------------------|
| asm_IdServicioMaterno | int | <input checked="" type="checkbox"/> |
| asm_IdAtencion | int | <input checked="" type="checkbox"/> |
| asm_IdSmi | nvarchar(9) | <input checked="" type="checkbox"/> |
| asm_Dato | nvarchar(60) | <input checked="" type="checkbox"/> |
| asm_Version | nvarchar(27) | <input checked="" type="checkbox"/> |
| P_Digi | nvarchar(255) | <input checked="" type="checkbox"/> |
| ate_Periodo | nvarchar(12) | <input checked="" type="checkbox"/> |
| ate_Mes | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Lote | nvarchar(6) | <input checked="" type="checkbox"/> |
| ate_Numero | nvarchar(24) | <input checked="" type="checkbox"/> |
| | | <input type="checkbox"/> |

Figura 44. Tabla SMI de la base de datos del modelo de sistema Web Scraping

Fuente: Elaboracion propia

Diagrama de base de datos

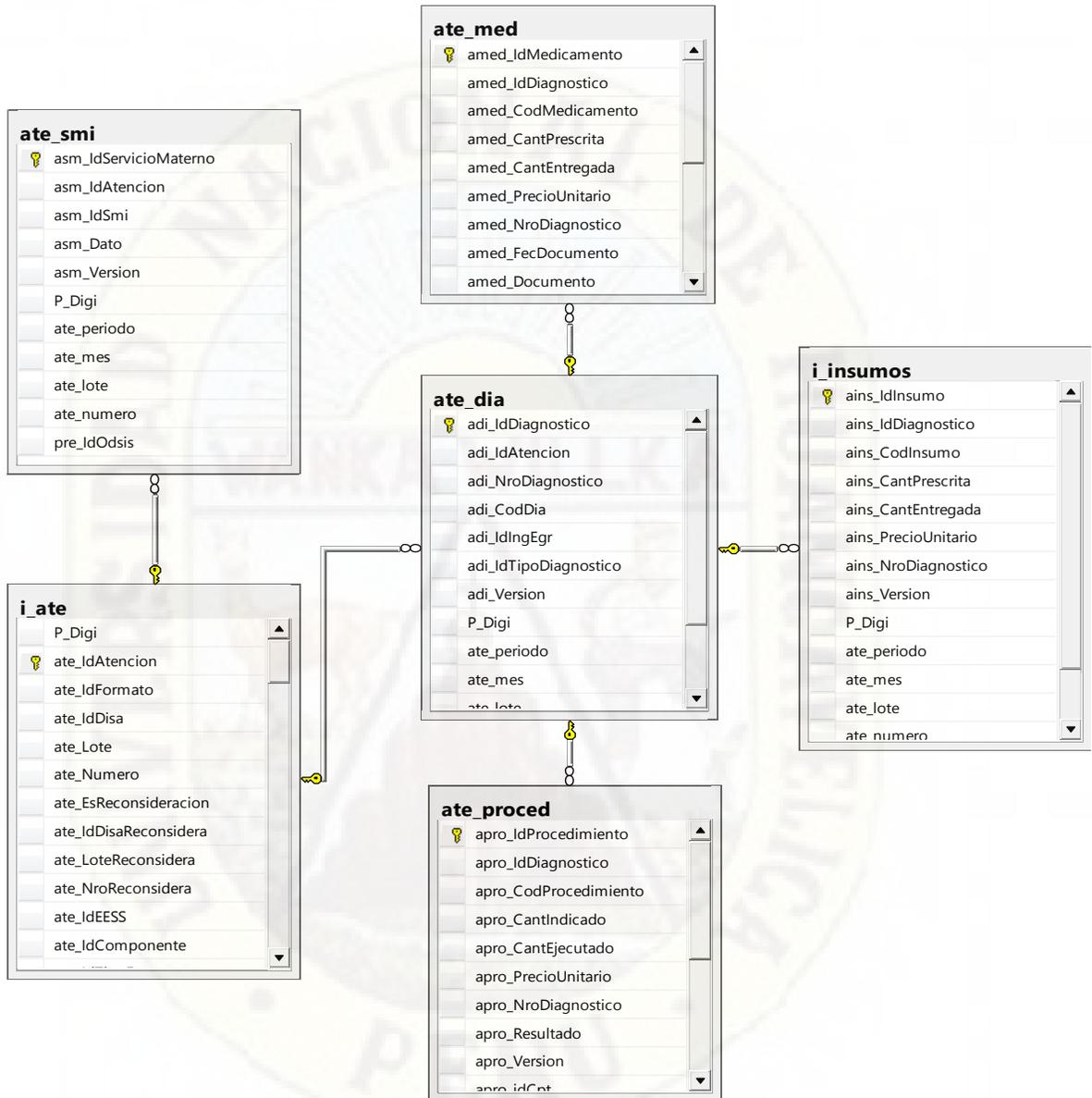


Figura 45. Diagrama de base de datos del modelo de sistema basado en Web Scraping

Ejecutando el Script de un indicador

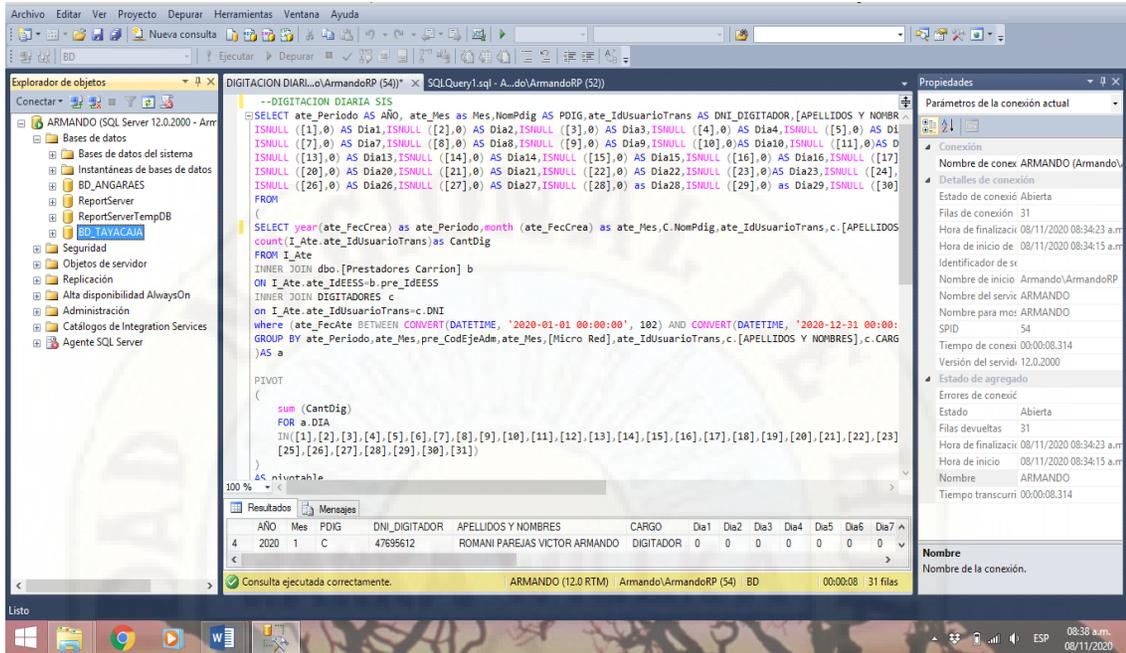


Figura 46. Ejecución de un Script de un indicador

Fuente: Elaboración propia

Resultado Final – El tablero estadístico en Excel



Figura 47. Diseño del tablero estadístico final

Fuente: Elaboración propia

Resultado Final - El tablero estadístico en Excel-reportes de asegurados

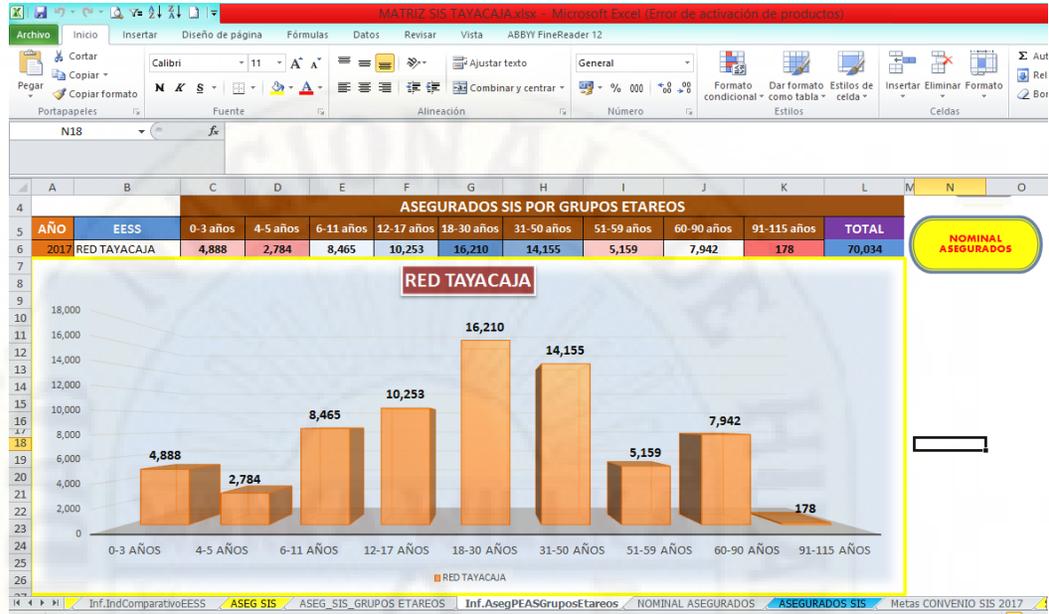


Figura 48. Tablero estadístico del reporte de asegurados

Fuente: Elaboración propia

Resultado Final – El tablero estadístico en Excel-seguimiento de niños

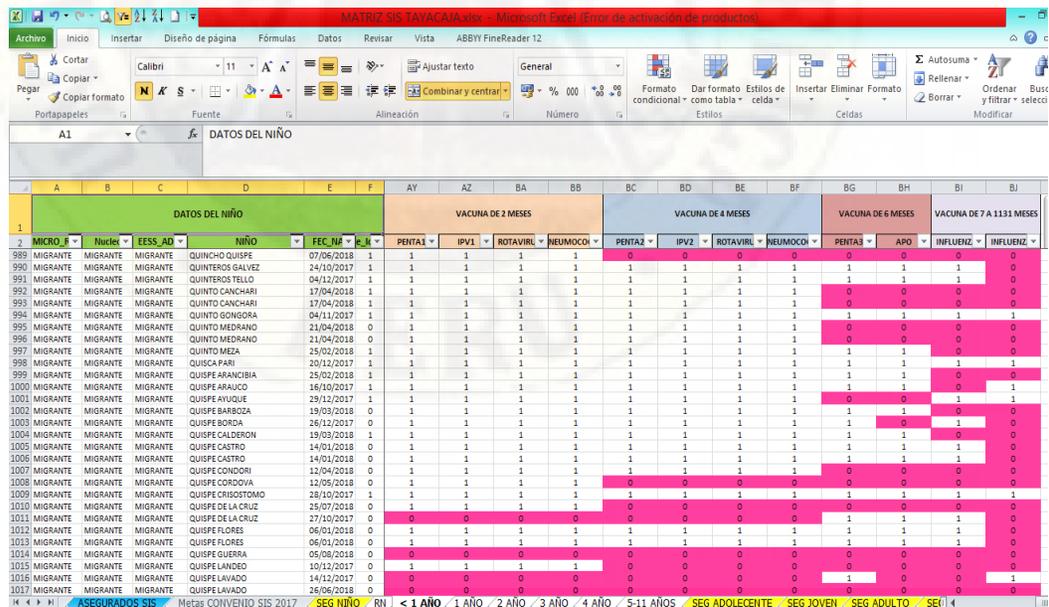


Figura 49. Tablero estadístico seguimiento de niños

Fuente: Elaboración propia

Script para el seguimiento de niños de 1 año

-- NOTA:

--EJECUTAR PRIMERO EL PAQUETE MATERNO DE SEGUIMIENTO

```
--SELECT * into ##I_Ate FROM I_Ate a where year(a.ate_fecate) in
(2017,2018)
-- cred
SELECT distinct b.*
into #CRED
FROM I_Ate b
INNER JOIN Ate_SMI SMI
ON b.ate_Lote+b.ate_numero+b. P_Digi =
SMI.ate_lote+SMI.ate_Numero+smi.P_Digi
WHERE (b.ate_IdServicio in ('001','002')) AND
(SMI.asm_IdSmi = '120')
AND (b.ate_Edades BETWEEN 10000 AND 11129)
-- vacuna NEUMOCOCO
SELECT distinct I.*
into #Neumococo
FROM I_Ate AS I
INNER JOIN Ate_SMI SMI
ON I.ate_Lote+i.ate_Numero+i.P_Digi=
SMI.ate_Lote+smi.ate_Numero+smi.P_Digi
WHERE (I.ate_Edades BETWEEN 10000 AND 11129)
AND (I.ate_IdServicio IN
('001','002','005','007','008','016','056','050','061','060','075'))
AND (SMI.asm_IdSmi in ('126'))
-- vacuna SPR
SELECT distinct I.*
into #SPR
FROM I_Ate AS I
INNER JOIN Ate_SMI SMI
ON I.ate_Lote+i.ate_Numero+i.P_Digi=
SMI.ate_Lote+smi.ate_Numero+smi.P_Digi
WHERE (I.ate_Edades BETWEEN 10000 AND 11129)
```

```

AND (I.ate_IdServicio IN
('001','002','005','007','008','016','056','050','061','060','075'))
AND (SMI.asm_IdSmi in ('125')) -- SPR
-- vacuna DPT
SELECT distinct I.*
into #DPT
FROM I_Ate AS I
INNER JOIN Ate_SMI SMI
ON I.ate_Lote+i.ate_Numero+i.P_Digi=
SMI.ate_Lote+smi.ate_Numero+smi.P_Digi
WHERE (I.ate_Edades BETWEEN 10000 AND 11129)
AND (I.ate_IdServicio IN
('001','002','005','007','008','016','056','050','061','060','075'))
AND (SMI.asm_IdSmi in ('117')) -- DPT
-- vacuna APO
SELECT distinct I.*
into #APO
FROM I_Ate AS I
INNER JOIN Ate_SMI SMI
ON I.ate_Lote+i.ate_Numero+i.P_Digi=
SMI.ate_Lote+smi.ate_Numero+smi.P_Digi
WHERE (I.ate_Edades BETWEEN 10000 AND 11129)
AND (I.ate_IdServicio IN
('001','002','005','007','008','016','056','050','061','060','075'))
AND (SMI.asm_IdSmi in ('313')) -- APO
-- MMN
SELECT distinct I.*, (m_medicamentos.med_Nombre+'
en '+m_medicamentos.med_FormaFarmaceutica+' Cant: '+convert (varchar (3),
MED.amed_CantEntregada)) as Medicamento
into #MMN
FROM I_Ate AS I
INNER JOIN Ate_Med AS MED
ON i.ate_Lote+i.ate_Numero+i.P_Digi =
MED.ate_Lote+med.ate_Numero+med.P_Digi
inner join bdsis_maestros.dbo.m_medicamentos on med.
amed_CodMedicamento=m_medicamentos.med_CodMed

```

```

WHERE (I.ate_IdServicio
IN('007','001','002','060','075')) AND (MED.amed_CodMedicamento
IN('S0001','03536','03519','28551','01759')
AND (I.ate_Edades BETWEEN 10000 AND 11129))

```

-- VISITA DOMICILIARIA

```

SELECT distinct I.*
into #VIS_DOMIC
FROM I_Ate AS I
INNER JOIN Ate_Dia AS DIA
ON i.ate_Lote+i.ate_Numero+i.P_Digi =
DIA.ate_Lote+DIA.ate_Numero+DIA.P_Digi
WHERE I.ate_IdServicio IN ('060','075')
AND I.ate_Edades BETWEEN 10000 AND 11129 -- de

```

un año

Union

```

SELECT distinct I.*
FROM I_Ate AS I
INNER JOIN Ate_Dia AS DIA
ON i.ate_Lote+i.ate_Numero+i.P_Digi =
DIA.ate_Lote+DIA.ate_Numero+DIA.P_Digi
WHERE dia.adi_CodDia iN('Z742','Z743')
AND I.ate_Edades BETWEEN 10000 AND 11129 -- de

```

un año

-- ESTIMULACIÓN TEMPRANA

```

SELECT distinct b.*
into #ET
FROM I_Ate b
INNER JOIN Ate_SMI SMI
ON b.ate_Lote+b.ate_numero+b. P_Digi =
SMI.ate_lote+SMI.ate_Numero+smi.P_Digi
WHERE b.ate_IdServicio in ('016')
AND B.ate_Edades BETWEEN 10000 AND 11129 --

```

de un año

union

```

SELECT distinct a.*
FROM I_Ate A INNER JOIN Ate_Proced D

```

```

ON
A.ate_Lote+A.ate_Numero+A.P_Digi=D.ate_Lote+D.ate_Numero+D.P_Digi
a.ate_IdEESS=pa.pre_IdEESS
inner join dbo. Prestador_Huancayo pa on
vinculadas
WHERE A.ate_IdServicio in ('001','016') -- para las
de ET
AND D. apro_CodProcedimiento IN ('9940105') -- cpt
de un año
AND A.ate_Edades BETWEEN 10000 AND 11129 --
-- TAMIZAJE VIF
SELECT distinct I.*
INTO #VIF
FROM I_Ate AS I
INNER JOIN Ate_Dia AS DIA
ON i.ate_Lote+i.ate_Numero+i.P_Digi =
DIA.ate_Lote+DIA.ate_Numero+DIA.P_Digi
un año
WHERE dia.adi_CodDia IN('Z133')
AND I.ate_Edades BETWEEN 10000 AND 11129 -- de
-- EXAMEN BUCAL
select distinct *
into #ExaBucal
from (
select I.*
FROM I_Ate AS I
INNER JOIN Ate_Proced proced
ON I.ate_Lote = proced.ate_lote
and I.ate_numero=proced.ate_Numero
and I.P_Digi=proced.P_Digi
WHERE (I.ate_IdServicio IN ('020'))
and (proced. apro_CodProcedimiento
in('99255','D0120','D0150')) -- EXAMEN BUCAL
AND I.ate_Edades BETWEEN 10000 AND
11129 -- de un año
) as a
-- FISIOTERAPIA

```

```

select distinct *
into #Fisioterapia
from (
    select I.*
    FROM I_Ate AS I
    INNER JOIN Ate_Proced proced
    ON I.ate_Lote = proced.ate_lote
    and I.ate_numero=proced.ate_Numero
    and I.P_Digi=proced.P_Digi
    WHERE (I.ate_IdServicio IN ('020'))
    and (proced. apro_CodProcedimiento
in('97782','D0140')) -- FISIOTERAPIA
    AND I.ate_Edades BETWEEN 10000 AND
11129 -- de un año
) as a
-- -- HB
select distinct a.*
into #HB
FROM I_Ate AS A
INNER JOIN Ate_Proced AS P
ON a.ate_Lote = P.ate_Lote
and a.ate_Numero=p.ate_Numero
and a.P_Digi=p.P_Digi
WHERE (P. apro_CodProcedimiento IN ('85018'))
AND (a.ate_Edades BETWEEN 10000 AND 11129)
-- resultado de HB
select distinct I.*, proced. apro_Resultado
into #ResultadoHB
FROM I_Ate AS I
INNER JOIN Ate_Proced proced
ON I.ate_Lote = proced.ate_lote
and I.ate_numero=proced.ate_Numero
and I.P_Digi=proced.P_Digi
WHERE (proced. apro_CodProcedimiento = '85018')
AND (i.ate_Edades BETWEEN 10000 AND 11129)

```

```

-- para la direccion del asegurado
--select distinct aus.af_i_DNI, aus.af_i_direccion
--into #AfiliadosDireccion
--from [SERVER\DAP]. BDODSIS_AFI_007_2019_04.DBO.I_AfiliacionAus aus
--where year(aus.af_i_fecnac) between year (GETDATE ())-2 and year (getdate ())
--union
--select aus.af_i_DNI, aus.AFI_DIRECCIONTITULAR
--from [SERVER\DAP]. BDODSIS_AFI_007_2019_04.DBO.I_AfiliacionAusSemi aus
--where year(aus.af_i_fecnac) between year (GETDATE ())-2 and year (getdate ())
--union
--select aus.af_i_DNI, aus.af_i_direcc
--from [SERVER\DAP]. BDODSIS_AFI_007_2019_04.DBO.I_AfiliacionOld aus
--where year(aus.af_i_fecnac) between year (GETDATE ())-2 and year (getdate ())
--union
--select aus.afa_Dni, aus.AFI_DIRECCIONTITULAR
--from [SERVER\DAP]. BDODSIS_AFI_007_2019_04.DBO.I_AfiliacionSemi aus
--where year(aus.afa_FecNac) between year (GETDATE ())-2 and year (getdate ())
--union
--select aus.af_i_DNI, aus.af_i_direccion
--from [SERVER\DAP]. BDODSIS_AFI_007_2019_04.DBO.I_AfiliacionSub aus
--where year(aus.af_i_fecnac) between year (GETDATE ())-2 and year (getdate ())
-- CONSULTA PRINCIPAL
SELECT distinct PA. N_RED, PA. [Micro Red], PA.pre_Nombre AS EESS--,
A.af_i_Correlativo AS COR
, ISNULL(A.af_i_DNI,0) AS DNI
, (A.af_i_appaterno +' '+ A.af_i_apmaterno +' '+ A.af_i_nombres) AS [APELLIDOS Y
NOMBRES], CONVERT (VARCHAR (10), A.af_i_fecnac,103) AS FEC_NAC
, (convert (char (2), dbo. fn_Edad_Años (A.af_i_fecnac, GETDATE ())) +' a') as Años
, (convert (char (2), dbo. fn_Edad_Años (A.af_i_fecnac, GETDATE ())) +' a '+ convert (char
(2),dbo.fn_Edad_Meses(A.af_i_fecnac, GETDATE()))+' m
'+convert(char(2),dbo.fn_Edad_AñosMesesDias(A.af_i_fecNac, GETDATE()))+' d')as 'Edad
Actual'
,
-- 12 MESES
--CRED

```

```

(SELECT      COUNT (*)
FROM #CRED b
WHERE (b.ate_ApePaterno = A.afi_appaterno) AND (b.ate_ApeMaterno =
A.afi_apmaterno)
AND (b.ate_FecNac = A.afi_fecnac) AND (b.ate_IdSexo = A.afi_idsexo)
AND (b.ate_Edades BETWEEN 10000 AND 10028)
) as '12mCRED',
-- VACUNA NEUMOCOCO
isnull ((SELECT COUNT (*)
FROM #Neumococo I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10000
AND 10028)
),0) AS '12mNEUMO',
-- VACUNA SPR
isnull ((SELECT COUNT (*)
FROM #SPR I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10000
AND 10028)
),0) AS '12mSPR',
-- MMN
(SELECT top 1 I. Medicamento
FROM #MMN I
WHERE      (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
= A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac) AND (I.ate_IdSexo = A.afi_idsexo)
and (I.ate_Edades BETWEEN 10000 AND 10028)
) AS '12mMMN',
-- VIF
isnull ((SELECT COUNT (*)
FROM #VIF I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10000
AND 10028)

```

```

),0) AS '12mVIF',
-- ET
isnull ((SELECT COUNT (*)
FROM #ET I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10000
AND 10028)
),0) AS '12mET',
-- VD
isnull ((SELECT COUNT (*)
FROM #VIS_DOMIC I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10000
AND 10028)
),0) AS '12mVD',
-- SB
isnull ((SELECT COUNT (*)
FROM #ExaBucal I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10000
AND 10028)
),0) AS '12mSB',
--Resultado HB
isnull ((select top 1 [dbo].[fn_Validar_Resultado_Hb](i.apro_Resultado) as
apro_Resultado
FROM #ResultadoHB I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10000
AND 10028)
order by i.ate_FecAte asc
),0) AS '12mHB',

-- Fecha de HB

```

```

isnull ((select top 1 (convert (varchar (10), i.ate_FecAte,103))
FROM #ResultadoHB I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10000
AND 10028)
order by i.ate_FecAte asc
),0) AS '12mFecHB',
-- 13 MESES
-- MMN
(SELECT top 1 I. Medicamento
FROM #MMN I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno
= A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac) AND (I.ate_IdSexo = A.afi_idsexo)
and (I.ate_Edades BETWEEN 10029 AND 10128)
) AS '13mMMN',
--14 MESES
-- CRED 2 meses
(SELECT COUNT (*)
FROM #CRED b
WHERE (b.ate_ApePaterno = A.afi_appaterno) AND (b.ate_ApeMaterno =
A.afi_apmaterno)
AND (b.ate_FecNac = A.afi_fecnac) AND (b.ate_IdSexo = A.afi_idsexo)
AND (b.ate_Edades BETWEEN 10129 AND 10228)
) as '14mCRED',
-- MMN
(SELECT top 1 I. Medicamento
FROM #MMN I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno
= A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac) AND (I.ate_IdSexo = A.afi_idsexo)
and (I.ate_Edades BETWEEN 10129 AND 10228)
) AS '14mMMN',
-- 15 MESES
-- MMN
(SELECT top 1 I.Medicamento
FROM #MMN I

```

```

WHERE      (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno
= A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac) AND (I.ate_IdSexo = A.afi_idsexo)
and (I.ate_Edades BETWEEN 10229 AND 10328)
) AS '15mMMN',
-- 16 meses
(SELECT      COUNT (*)
FROM #CRED b
WHERE (b.ate_ApePaterno = A.afi_appaterno) AND (b.ate_ApeMaterno =
A.afi_apmaterno)
AND (b.ate_FecNac = A.afi_fecnac) AND (b.ate_IdSexo = A.afi_idsexo)
AND (b.ate_Edades BETWEEN 10329 AND 10428)
) as '16mCRED',
-- MMN
(SELECT top 1 I. Medicamento
FROM #MMN I
WHERE      (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno
= A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac) AND (I.ate_IdSexo = A.afi_idsexo)
and (I.ate_Edades BETWEEN 10329 AND 10428)
) AS '16mMMN',
-- 17 MESES
-- MMN
(SELECT top 1 I. Medicamento
FROM #MMN I
WHERE      (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno
= A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac) AND (I.ate_IdSexo = A.afi_idsexo)
and (I.ate_Edades BETWEEN 10429 AND 10528)
) AS '17mMMN',
-- 18 MESES
(SELECT      COUNT (*)
FROM #CRED b
WHERE (b.ate_ApePaterno = A.afi_appaterno) AND (b.ate_ApeMaterno =
A.afi_apmaterno)
AND (b.ate_FecNac = A.afi_fecnac) AND (b.ate_IdSexo = A.afi_idsexo)

```

```

AND (b.ate_Edades BETWEEN 10529 AND 10628)
) as '18mCRED',
-- VAC DPT
(SELECT      COUNT (*)
FROM #DPT b
WHERE (b.ate_ApePaterno = A.af_i_appaterno) AND (b.ate_ApeMaterno =
A.af_i_apmaterno)
AND (b.ate_FecNac = A.af_i_fecnac) AND (b.ate_IdSexo = A.af_i_idsexo)
AND (b.ate_Edades BETWEEN 10529 AND 10628)
) as '18mDPT',
-- VAC APO
(SELECT      COUNT (*)
FROM #APO b
WHERE (b.ate_ApePaterno = A.af_i_appaterno) AND (b.ate_ApeMaterno =
A.af_i_apmaterno)
AND (b.ate_FecNac = A.af_i_fecnac) AND (b.ate_IdSexo = A.af_i_idsexo)
AND (b.ate_Edades BETWEEN 10529 AND 10628)
) as '18mAPO',
-- VACUNA SPR
isnull ((SELECT COUNT (*)
FROM #SPR I
WHERE (I.ate_ApePaterno = A.af_i_appaterno) AND (I.ate_ApeMaterno =
A.af_i_apmaterno) AND (I.ate_FecNac = A.af_i_fecnac)
AND (I.ate_IdSexo = A.af_i_idsexo) AND (I.ate_Edades BETWEEN 10529
AND 10628)
),0) AS '18mSPR',
-- ET
isnull ((SELECT COUNT (*)
FROM #ET I
WHERE (I.ate_ApePaterno = A.af_i_appaterno) AND (I.ate_ApeMaterno =
A.af_i_apmaterno) AND (I.ate_FecNac = A.af_i_fecnac)
AND (I.ate_IdSexo = A.af_i_idsexo) AND (I.ate_Edades BETWEEN 10529
AND 10628)
),0) AS '18mET',
-- VD
isnull ((SELECT COUNT (*)

```

```

FROM #VIS_DOMIC I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10529
AND 10628)
),0) AS '18mVD',
-- SB
isnull ((SELECT COUNT (*)
FROM #ExaBucal I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10529
AND 10628)
),0) AS '18mSB',
--Resultado HB
isnull ((select top 1 [dbo]. [fn_Validar_Resultado_Hb] (i. apro_Resultado) as
apro_Resultado
FROM #ResultadoHB I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10529
AND 10628)
order by i.ate_FecAte asc
),0) AS '18mHB',
-- Fecha de HB
isnull ((select top 1 (convert (varchar (10), i.ate_FecAte,103))
FROM #ResultadoHB I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10529
AND 10628)
order by i.ate_FecAte asc
),0) AS '18mFecHB',
-- 20 MESES
(SELECT COUNT (*)
FROM #CRED b

```

```

WHERE (b.ate_ApePaterno = A.afi_appaterno) AND (b.ate_ApeMaterno =
A.afi_apmaterno)
AND (b.ate_FecNac = A.afi_fecnac) AND (b.ate_IdSexo = A.afi_idsexo)
AND (b.ate_Edades BETWEEN 10729 AND 10828)
) as '20mCRED',
-- MMN
(SELECT top 1 I. Medicamento
FROM #MMN I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno
= A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac) AND (I.ate_IdSexo = A.afi_idsexo)
and (I.ate_Edades BETWEEN 10729 AND 10828)
) AS '20mMMN',

-- 21 MESES
-- ET
isnull ((SELECT COUNT (*)
FROM #ET I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10829
AND 10928)
),0) AS '21mET',
-- VD
isnull ((SELECT COUNT (*)
FROM #VIF I
WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac)
AND (I.ate_IdSexo = A.afi_idsexo) AND (I.ate_Edades BETWEEN 10829
AND 10928)
),0) AS '21mVIF',

-- 22 MESES
(SELECT COUNT (*)
FROM #CRED b
WHERE (b.ate_ApePaterno = A.afi_appaterno) AND (b.ate_ApeMaterno =
A.afi_apmaterno)
AND (b.ate_FecNac = A.afi_fecnac) AND (b.ate_IdSexo = A.afi_idsexo)
AND (b.ate_Edades BETWEEN 10929 AND 11028)

```

```

        ) as '22mCRED',
-- Fecha de ultimo HB
        isnull ((SELECT top 1 Convert (varchar (10), ate_FecAte,103)
        FROM #ResultadoHB AS I
        WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac) AND (I.ate_IdSexo = A.afi_idsexo)
        order by I.ate_FecAte Desc
        ),'No tiene') as 'Fecha Último Dosaje de HB',
-- HB_ con resultado
        isnull ((SELECT top 1 apro_Resultado
        FROM #ResultadoHB AS I
        WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac) AND (I.ate_IdSexo = A.afi_idsexo)
        order by I.ate_FecAte Desc
        ),'No tiene') as 'HB_UltResult',
-- Ultimo EESS de Atención el presente Año
        (SELECT TOP 1 pa.pre_Nombre+' Serv:'+convert (char (3), i.ate_IdServicio) +'
FecAte:'+convert (varchar (10), i.ate_FecAte,103)
        FROM I_Ate I inner join dbo. Prestador_Huancayo pa on i.ate_IdEESS=pa.pre_IdEESS
        WHERE (I.ate_ApePaterno = A.afi_appaterno) AND (I.ate_ApeMaterno =
A.afi_apmaterno) AND (I.ate_FecNac = A.afi_fecnac) AND (I.ate_IdSexo = A.afi_idsexo)
        ORDER BY i.ate_FecAte DESC
        ) AS 'Ult.EESS Atención'
--, -- ESTADO DE LA AFILIACIÓN
--a.afi_estado AS 'Estado Afiliación',
---- DIRECCIÓN
--isnull ((SELECT TOP 1 AUS.afi_direccion
--FROM #AfiliadosDireccion AUS
--WHERE (AUS.afi_DNI = A.afi_DNI COLLATE Modern_Spanish_CS_AS)
--),' ') AS 'DIRECCION'
into #Seg_1año
FROM dbo.Prestador_Huancayo AS PA
INNER JOIN dbo. Afiliados_DISA AS A
ON PA.pre_IdEESS COLLATE SQL_Latin1_General_CP1_CI_AS =
A.afi_idEESSAdscripcion
WHERE (A.afi_fecnac BETWEEN CONVERT (DATETIME, '2017-01-01 00:00:00', 102)

```

```

and CONVERT (DATETIME, '2018-06-30 00:00:00', 102)) -- SE ANALIZA 11 MESES
ATRAS AL MES ACTUAL
ORDER BY PA. N_RED, [Micro Red], EESS, [APELLIDOS Y NOMBRES] asc
--- Consulta Final con detalle
select distinct N_RED, [Micro Red], EESS, DNI, [APELLIDOS Y NOMBRES]
, FEC_NAC, Años, [Edad Actual],
case when [12mCRED] =0 then 'No Tiene' else convert(varchar(10),[12mCRED]) end
[12mCRED],
case when [12mNEUMO] =0 then 'No Tiene' else convert(varchar(10),[12mNEUMO]) end
[12mNEUMO],
case when [12mSPR] =0 then 'No Tiene' else convert(varchar(10),[12mSPR]) end [12mSPR],
isnull([12mMMN],'No Tiene')[12mMMN],
case when [12mVIF] =0 then 'No Tiene' else convert(varchar(10),[12mVIF]) end [12mVIF],
case when [12mET] =0 then 'No Tiene' else convert(varchar(10),[12mET]) end [12mET],
case when [12mVD] =0 then 'No Tiene' else convert(varchar(10),[12mVD]) end [12mVD],
case when [12mSB] =0 then 'No Tiene' else convert(varchar(10),[12mSB]) end [12mSB],
isnull([12mHB],'No Tiene')[12mHB],
isnull([12mFecHB],'No Tiene')[12mFecHB],
isnull([13mMMN],'No Tiene')[13mMMN],
case when [14mCRED] =0 then 'No Tiene' else convert(varchar(10),[14mCRED]) end
[14mCRED],
isnull([14mMMN],'No Tiene')[14mMMN],
isnull([15mMMN],'No Tiene')[15mMMN],
case when [16mCRED] =0 then 'No Tiene' else convert(varchar(10),[16mCRED]) end
[16mCRED],
isnull([16mMMN],'No Tiene')[16mMMN],
isnull([17mMMN],'No Tiene')[17mMMN],
case when [18mCRED] =0 then 'No Tiene' else convert(varchar(10),[18mCRED]) end
[18mCRED],
case when [18mDPT] =0 then 'No Tiene' else convert(varchar(10),[18mDPT]) end
[18mDPT],
case when [18mAPO] =0 then 'No Tiene' else convert(varchar(10),[18mAPO]) end
[18mAPO],
case when [18mSPR] =0 then 'No Tiene' else convert(varchar(10),[18mSPR]) end [18mSPR],
case when [18mET] =0 then 'No Tiene' else convert(varchar(10),[18mET]) end [18mET],
case when [18mVD] =0 then 'No Tiene' else convert(varchar(10),[18mVD]) end [18mVD],
case when [18mSB] =0 then 'No Tiene' else convert(varchar(10),[18mSB]) end [18mSB],

```

```

isnull([18mHB],'No Tiene') [18mHB],
isnull([18mFecHB],'No Tiene') [18mFecHB],
case when [20mCRED] =0 then 'No Tiene' else convert (varchar (10), [20mCRED]) end
[20mCRED],
isnull([20mMMN],'No Tiene') [20mMMN],
case when [21mET] =0 then 'No Tiene' else convert (varchar (10), [21mET]) end [21mET],
case when [21mVIF] =0 then 'No Tiene' else convert (varchar (10), [21mVIF]) end
[21mVIF],
case when [22mCRED] =0 then 'No Tiene' else convert (varchar (10), [22mCRED]) end
[22mCRED],
isnull ([Fecha Último Dosaje de HB],'No Tiene') [Fecha Último Dosaje de HB],
isnull (HB_UltResult,'No Tiene') HB_UltResult,
isnull ([Ult.EESS Atención],'No Tiene') [Ult.EESS Atención]
--, case when [Estado Afiliación]=0 then 'ACTIVO' else 'INACTIVO' end [Estado
Afiliación],
--isnull([DIRECCION],' ') [DIRECCION]
from #Seg_1año
DROP TABLE #CRED
DROP TABLE #HB
DROP TABLE #MMN
DROP TABLE #Neumococo
DROP TABLE #ResultadoHB
DROP TABLE #SPR
drop table #Seg_1año
drop table #DPT
drop table #APO
drop table #ET
drop table #ExaBucal
drop table #Fisioterapia
drop table #VIF
drop table #VIS_DOMIC
GO

```