

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

**WEBCRAWLER DE PROCESOS DE SUBASTA INVERSA
ELECTRÓNICA DEL PORTAL DE COMPRAS PÚBLICAS
DEL ECUADOR**

Sebastián Alejandro Taraguay Silva

Ingeniería en Ciencias de la Computación

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniero en Ciencias de la Computación

Quito, 20 de mayo de 2021

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

**HOJA DE CALIFICACIÓN
DE TRABAJO DE FIN DE CARRERA**

**WEBCRAWLER DE PROCESOS DE SUBASTA INVERSA
ELECTRÓNICA DEL PORTAL DE COMPRAS PÚBLICAS
DEL ECUADOR**

Sebastián Alejandro Taraguay Silva

Nombre del profesor, Título académico

Daniel Riofrío, PhD.

Quito, 20 de mayo de 2021

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Sebastián Alejandro Taraguay Silva

Código: 139248

Cédula de identidad: 1723028096

Lugar y fecha: Quito, 20 de mayo de 2021

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETHeses>.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETHeses>.

RESUMEN

Dentro de los Procesos de Contratación Pública, el mecanismo de subasta inversa electrónica ha conformado una parte significativa del total de contrataciones públicas, llegando en el año de 2019 según la SERCOP al 34.1% del total de procesos de contratación pública del país. Todos los registros detallados de este tipo de contratación se encuentran publicados en el portal web de la SERCOP, sin embargo, existen ciertos detalles que no se pueden visualizar en ciertas etapas del proceso; además, la información dentro del portal web no se encuentra presentada de una manera clara y sencilla para los usuarios finales, debido a esto se dificulta la transparencia de los procesos. La motivación principal de este proyecto es facilitar y presentar de manera clara y sencilla los detalles más relevantes de cada proceso de subasta inversa electrónica con el fin de servir de base a posibles trabajos futuros que auditen este tipo de contrataciones en busca de posibles irregularidades. A través de la implementación de un Web Crawler escrito en Python y con ayuda del módulo Selenium, se pretende cumplir con el objetivo principal.

Palabras clave: contratación pública, subasta inversa electrónica, web crawler, corrupción, Python, base de datos

ABSTRACT

Within Public Procurement Processes, the electronic reverse auction mechanism has made up a significant part of the total public procurement, reaching 34.1% of the total public procurement processes in Ecuador in 2019 according to SERCOP. All the detailed records of this type of contracting are published on the SERCOP web portal, however, there are certain details that cannot be viewed at certain stages of the process. In addition, the information within the web portal is not presented in a clear and simple way for end users, due to this, transparency is an issue. The main motivation of this project is to facilitate and present in a clear and simple way the most relevant details of each electronic reverse auction process in order to serve as a basis for possible future work that audits this type of contracting in search of possible irregularities. Through the implementation of a Web Crawler written in Python and with the help of the Selenium module, it is intended to meet the main objective.

Key words: public procurement, electronic reverse auction, web crawler, corruption, Python, database.

TABLA DE CONTENIDO

Introducción	10
Antecedentes	10
Objetivos	11
Objetivo General.....	11
Objetivo Específicos	11
Desarrollo del Tema.....	12
Subasta inversa electrónica.....	12
El sistema oficial de contratación pública	13
Estructura de navegación en el portal web de la SERCOP	16
Web Crawler	22
Módulo de navegación.....	23
Módulo de extracción	28
Módulo de almacenamiento	31
Módulo de utilidades	32
Tecnologías utilizadas	33
Mejoras propuestas.....	33
Resultados.....	34
Conclusiones	36
Referencias bibliográficas.....	38
Referencias.....	38
Anexo A: Documentación del programa	39

ÍNDICE DE IMÁGENES

Imagen 1: Página de búsqueda de procesos de contratación.	14
Imagen 2: Tabla de resultado de procesos subasta inversa electrónica	14
Imagen 3: Página con los detalles de cada proceso.	15
Imagen 4: Códigos únicos de cada proceso.	17
Imagen 5: Información relevante para ser extraída.....	17
Imagen 6: Información relevante para ser extraída.....	18
Imagen 7: Información relevante para ser extraída.....	18
Imagen 8: Información extra del proceso	19
Imagen 9: Lista de proveedores invitados a participar	19
Imagen 10: Información relevante del resumen de contrato.....	20
Imagen 11: Información a extraer en el nivel Index	25
Imagen 12: Página correspondiente al nivel 1 de navegación	26
Imagen 13: Página de invitaciones a proveedores	27
Imagen 14: Página de información esencial del contrato	28
Imagen 15: Información extraída del nivel Index	29
Imagen 16: Información extraída del nivel, descripción del proceso	29
Imagen 17: Información extraída del nivel 1, fechas de control del proceso	30
Imagen 18: Información extraída del nivel 2, Invitaciones a proveedores.	30
Imagen 19: Información extraída del nivel 2, información esencial del contrato	30
Imagen 20: Esquema de la base de datos	32

ÍNDICE DE ILUSTRACIONES

Ilustración 1: Flujo de proceso SIE.....	16
Ilustración 2: Proceso cíclico de obtención de procesos.....	20
Ilustración 3: Flowchart del Web Crawler.....	21
Ilustración 4: Módulos del Web Crawler.....	23

ÍNDICE DE TABLAS

Tabla 1: Funciones del módulo de navegación.....	24
Tabla 2: Funciones en el módulo de extracción.....	28
Tabla 3: Funciones del módulo almacenamiento.....	31
Tabla 4: Tecnologías utilizadas para el desarrollo del proyecto.....	33
Tabla 5: Cantidad de datos almacenados en la BD.....	35

INTRODUCCIÓN

La subasta inversa electrónica hace referencia a los procedimientos dinámicos que son utilizados en la adquisición de bienes y servicios normalizados, en donde los proveedores pujan hacia la baja del precio ofertado. En el año 2019 este tipo de procesos de contratación pública abarcó el 34.1% del total. A pesar de que la información se encuentra de manera pública en el portal web de la SERCOP, existen ciertos detalles que no se pueden visualizar en ciertas etapas del proceso por lo cual se puede algún tipo de caso de corrupción. De ahí nace la necesidad de comprender la estructura y funcionamiento del portal web de consulta sobre los procesos de contratación pública para posteriormente automatizar la obtención y almacenamiento de la información relevante, mediante la implementación de un WEBCRAWLER(bot) desarrollado en base a Python3 y la librería Selenium el cual de manera automática se encargará de consultar en el portal web los procesos de Subasta Inversa Electrónica y extraer la información que posteriormente será almacenada de manera local en una base de datos manejada por MySQL con el fin de a futuro poder analizar los registros obtenidos en busca de irregularidades en cuanto a temas de contratación pública.

ANTECEDENTES

En el Ecuador actualmente existen 35 tipos de procesos de contratación pública, de los cuales según el *Análisis Anual de la Contratación Pública – 2019* publicado por la SERCOP - SOCE el 34.1% del total corresponde al método de subasta inversa electrónica es decir un monto correspondiente a 1.626,8 millones de dólares. Si bien el SOCE tiene como obligación facilitar al público en general la información acerca de los procedimientos de contratación pública, (salvo ciertas excepciones las cuales deben ser debidamente justificadas por parte del gobierno central), en el caso de Subasta Inversa Electrónica luego que el procedimiento pasa a los estados En *Recepción* o *Finalizados* no es posible visualizar a detalle la calificación

obtenida de cada proveedor participante ni el historial de puja o negociación del proceso en específico, lo cual puede llevar a casos de corrupción en la adjudicación de procesos debido a la falta de transparencia total. Por ello nace la necesidad de automatizar la recopilación y organización de los procesos publicados en el portal web de la SERCOP facilitando a posibles investigaciones futuras.

OBJETIVOS

Objetivo General

- Implementar un webcrawler que recopile la información de los procesos de subasta inversa electrónica de dominio público, alojados en el portal web de la SERCOP

Objetivo Específicos

- Analizar el webcrawler antecesor al presente trabajo, el cual es especializado en los procesos de contratación de emergencia.
- Comprender los datos que intervienen en los procesos de subasta inversa electrónica.
- Implementar un prototipo del webcrawler utilizando Python y la librería Selenium y sentar una base para extensiones a futuro del proyecto
- Diseñar un webcrawler basado en conceptos de reutilización de código y enfocarlo en los procesos de subasta inversa electrónica bajo tres principales actores: entidades contratantes, empresas proveedoras y etapas de contratación.

DESARROLLO DEL TEMA

Subasta inversa electrónica

La contratación mediante el procedimiento de subasta inversa se rige a los lineamientos de la Ley Orgánica del Servicio Nacional de Contratación Pública (LOSNCP) en la misma se señala en la sección 2, artículo 47 que:

“Para la adquisición de bienes y servicios normalizados que no consten en el catálogo electrónico, las Entidades Contratantes deberán realizar subastas inversas en las cuales los proveedores de bienes y servicios equivalentes pujan hacia la baja el precio ofertado, en acto público o por medios electrónicos a través del Portal de COMPRAS PÚBLICAS” (SERCOP, LEY ORGANICA DEL SISTEMA NACIONAL DE, 2018).

Se entiende como bienes y servicios normalizados “Aquellos cuyas características o especificaciones técnicas han sido estandarizadas u homologadas por la entidad contratante; y, en consecuencia, dichas características o especificaciones son homogéneas y comparables en igualdad de condiciones” (SERCOP, 2016).

Todo el proceso de subasta inversa electrónica es llevado a cabo en el portal de compras públicas de la SERCOP y se entiende como un procedimiento dinámico que es utilizado en la adquisición de bienes y servicios normalizados siempre y cuando su cuantía supere el monto equivalente al 0.0000002 del Presupuesto Inicial del Estado. La principal diferencia de este tipo de subasta inversa con el mecanismo de una subasta corriente radica en que los ofertantes pujan hacia la baja del precio ofertado inicial, siendo el ganador aquel que haya enviado la propuesta con el costo económico más bajo. Según los lineamientos de la SERCOP cada nueva puja tiene un tiempo de duración entre 15 a 60 minutos y su variación se recomienda entre el 0.5% y el 2% del precio referencial. En caso de que se presente una única oferta técnica calificada, no se continuará con la subasta y en su lugar se procederá a realizar la sesión única

de negociación entre el contratante y el oferente, el objetivo de la sesión será el mejorar la oferta económica y en caso de llegar a un acuerdo, se procederá a realizar la contratación con el único oferente. (SERCOP, 2016)

El sistema oficial de contratación pública

Las páginas web de la SERCOP, en donde se encuentran los registros de los procesos de contratación pública son muy poco intuitivas y amigables con el usuario, debido a esto puede resultar complejo para un usuario no experimentado la consulta y obtención de la información de los procesos. La página principal, como se muestra en la figura 1, cuenta con un recuadro a la derecha en la cual se encuentra la opción de búsqueda de procesos, al seleccionar dicha opción el usuario es redirigido a una segunda página web en la cual se puede consultar todo tipo de procesos en base a diferentes filtros como: palabras clave, entidad contratante, fechas de publicación y tipo de contratación. En este caso todos los registros serán obtenidos en base al filtro de tipo de contratación estableciendo el valor en *subasta inversa electrónica*. Es importante señalar que el sistema cuenta con un sistema de verificación CAPTCHA definido como:

“Completely Automated Public Turing test to tell Computers and Humans Apart”, que literalmente significa Test Público Completamente Automatizado de Turing para Diferenciar a los Seres Humanos de las Computadoras. Por lo tanto, viene a ser un Test de Turing inverso, en el que son los humanos los que tienen que demostrar con esta tecnología que son humanos y no máquinas” (FERNÁNDEZ, 2018).

Cada proceso listado en la tabla cuenta con los siguientes campos como valores de columna: código, entidad contratante, objeto del proceso, estado del proceso, provincia/cantón, presupuesto referencial total y fecha de publicación. El identificador único de cada proceso listado en la columna de *código* hace la función de un enlace el cual redirige al usuario a una página externa la cual contiene información detallada y relevante del proceso de subasta inversa electrónica tal como: autoridades a cargo, miembros de la comisión técnica, vigencia de la oferta, estado del proceso, valoración mínima de la puja, fechas relacionadas al proceso, archivos de respaldo y demás información. Cabe señalar que en la fase precontractual se añaden dos enlaces extras los cuales redirigen a otra página en la cual se encuentran preguntas y aclaraciones sobre el proceso y un segundo enlace conduce a otra página en la cual se lista en un formato de tabla los proveedores que fueron invitados a participar. Si el proceso fue finalizado se añade un enlace el cual contiene todo el resumen del proceso (ver imagen 3), todo el flujo de estados del proceso es representado en la ilustración 1.

Descripción del Proceso de Contratación

Entidad: HOSPITAL GENERAL MANTA

Objeto de Proceso: ADQUISICIÓN DE 17 ITEMS DE FARMACOS PARA EL HOSPITAL GENERAL MANTA

Código: SIE-IESSMANTA-48-20

Tipo Compra: Bien

Presupuesto Referencial Total (Sin Iva): USD 193,723.51

Tipo de Contratación: Subasta Inversa Electrónica

Cédula / Núm. Identificación	Apellidos y Nombres	Cargo	Origen
1308785060	COELLO VILLACIS JAIME FERNANDO	Máxima Autoridad Institucional	NACIONAL
1308785060	COELLO VILLACIS JAIME FERNANDO	Máxima Autoridad Responsable de la Gestión Administrativa - Financiera	NACIONAL
1308785060	COELLO VILLACIS JAIME FERNANDO	Autoridad Ordenadora de Gasto	NACIONAL

Comisión Técnica: Si

Cédula / Núm. Identificación	Apellidos y Nombres	Función en la Comisión Técnica
1308655347	VERA ALAVA FABRICIO JAVIER	Profesional designado por la máxima autoridad, quien lo presidirá
0603893058	COBA ROBALINO SONIA ELIZABETH	Titular del área requirente
1309450300	SOSA ALVARADO ALDRIN DIOGENES	Profesional afin al objeto de la contratación designado por la máxima

Imagen 3: Página con los detalles de cada proceso.



Ilustración 1: Flujo de proceso SIE

Estructura de navegación en el portal web de la SERCOP

El flujo de navegación a llevarse a cabo por el Web Crawler es de manera secuencial, comenzando en la página de búsqueda de procesos la cual presenta una tabla con los veinte procesos más recientes; en cada nivel de navegación se procederá a recolectar la información más relevante para ser almacenada en la base de datos correspondiente. Con el objetivo de evitar posibles bloqueos, en caso de que sea detectado el Web Crawler en el sistema, en cada nivel de navegación se recolectará de manera independiente la información.

El inicio del Web Crawler será a partir de la página de búsqueda (nivel Index) en la cual se presenta la tabla de procesos y su primer elemento de cada fila es un código único (ver imagen 4) que hace la función de enlace a otra página web (nivel 1) en la cual se detalla la información del proceso de contratación. Posteriormente el Web Crawler se encargará de la recolección y almacenamiento de los datos más relevantes, mediante sus módulos segundo y tercero.

Código	Entidad Contratante	Objeto del Proceso	Estado del Proceso	Provincia/Cantón	Presupuesto Referencial Total (sin iva)	Fecha de Publicación	Opciones
SIE-AGUAPAS-001-2021	EMPRESA PUBLICA DE AGUA ALCANTARILLADO Y ASEO DE PASAJE AGUAPAS EP	ADQUISICION DE INSUMO PARA TRATAMIENTO DE AGUA SUBTERRANEA EN SISTEMA DE BOMBEO Y POZO PROFUNDO DEL CANTON PASAJE	Preguntas, Respuestas y Aclaraciones	EL ORO / PASAJE	NO DISPONIBLE	2021-02-25 16:00:00	
SIE-ANDEC-004-21	AGENCIAS NACIONALES DEL ECUADOR SOCIEDAD ANONIMA (A.N.D.E.C.)	SERVICIO DE TRANSPORTE PARA LOS TRABAJADORES DEL CENTRO DE DISTRIBUCIÓN QUITO	Preguntas, Respuestas y Aclaraciones	GUAYAS / GUAYAQUIL	NO DISPONIBLE	2021-02-25 16:00:00	
SIE-CNELMIG-001-21	EMPRESA ELÉCTRICA PÚBLICA ESTRATÉGICA CORPORACIÓN NACIONAL DE ELECTRICIDAD CNEL EP	MLG MINTO ADQUISICION DE HERRAJES PARA EL MANTENIMIENTO DEL SISTEMA GD	Preguntas, Respuestas y Aclaraciones	GUAYAS / GUAYAQUIL	NO DISPONIBLE	2021-02-25 20:00:00	
SIE-EERSA-SDS-05-21	EMPRESA ELECTRICA RIOBAMBA S.A.	CONTRATACION DEL SERVICIO PARA TOMA DE PRUEBAS PCR, CUANTITATIVAS Y DE ANTIGENO PARA DETECCIÓN DE COVID-19	Preguntas, Respuestas y Aclaraciones	CHIMBORAZO / RIOBAMBA	NO DISPONIBLE	2021-02-25 20:00:00	
SIE-FMCIRES-003-2021	Empresa Pública Metropolitana de Gestión Integral de Residuos Sólidos	ADQUISICIÓN E INSTALACIÓN DE UN SISTEMA AIREACIÓN Y BOMBEO PARA EL TRATAMIENTO DE LIXIVIADOS DEL RELLENO SANITARIO DE QUITO	Preguntas, Respuestas y Aclaraciones	PICHINCHA / QUITO	NO DISPONIBLE	2021-02-25 17:00:00	
SIE-EPCONST-001-2021	EMPRESA PUBLICA DE OBRAS PUBLICAS, VIALIDAD, VIVIENDA, SEGURIDAD CIUDADANA Y DE ASEO DE SANTO DOMINGO EP-CONST	SERVICIO DE SEGURIDAD Y VIGILANCIA PRIVADA PARA TODAS LAS DEPENDENCIAS E INSTALACIONES DE LA EP-CONST	Preguntas, Respuestas y Aclaraciones	SANTO DOMINGO DE LOS TSACHILAS / SANTO DOMINGO	NO DISPONIBLE	2021-02-25 17:00:00	
SIE-EPMAPS-001-2021	Empresa Pública Metropolitana de Agua Potable y saneamiento	CARACTERIZACIONES DE DESCARGAS LÍQUIDAS Y RUIDO AMBIENTAL EN INSTALACIONES DE LA EPMAPS 2021	Preguntas, Respuestas y Aclaraciones	PICHINCHA / QUITO	NO DISPONIBLE	2021-02-25 18:00:00	
SIE-GADBOL-004-2020	Municipio de Bolívar	ADQUISICIÓN DE STICKERS ADHESIVOS PARA LA REVISIÓN TÉCNICA VEHICULAR DE VEHÍCULOS Y MOTOS	Preguntas, Respuestas y Aclaraciones	MANABI / BOLIVAR	NO DISPONIBLE	2021-02-25 18:00:00	
SIE-GADMCH-001-2021	MUNICIPIO DE CHONE	ADQUISICIÓN DE COMBUSTIBLES DIESEL, ECOPAIS Y SUPER PARA LOS VEHÍCULOS, MAQUINARIAS Y EQUIPOS DEL GOBIERNO AUTÓNOMO DESCENTRALIZADO DEL CANTÓN CHONE	Preguntas, Respuestas y Aclaraciones	MANABI / CHONE	NO DISPONIBLE	2021-02-25 20:00:00	
SIE-GADMCO-001-2021	CAD MUNICIPAL DE QUININDE	ALQUILER DE VEHICULOS QUE DARÁN EL SERVICIO DE RECOLECCIÓN DE LOS DESECHOS EN LOS DIFERENTES BARRIOS DE LA ZONA URBANA Y RURAL DEL CANTON QUININDE	Preguntas, Respuestas y Aclaraciones	ESMERALDAS / QUININDE	NO DISPONIBLE	2021-02-25 18:30:00	
SIE-GADMCS-06-21	Municipio del Cantón Samborondón	ADQUISICION DE UNA 1 ESTACION DE REFERENCIA CON RECEPTOR DE DOBLE FRECUENCIA Y ANTENA PARA SUMINISTRAR DATOS EN TIEMPO REAL Y REALIZAR CAPTURA DE DATOS SOLICITADO POR LA DIRECCIÓN DE AVALUOS Y CATASTRO	Preguntas, Respuestas y Aclaraciones	GUAYAS / SAMBORONDON	NO DISPONIBLE	2021-02-25 16:00:00	

Imagen 4: Códigos únicos de cada proceso.

Según el estado en el que se encuentre el proceso existirá información extra dentro de las fases precontractuales y contractuales. Por ello la función del Web Crawler será detectar el estado y en base al mismo obtener la información disponible necesaria.

Una vez analizada la pantalla de descripción del proceso (nivel 1) se estableció como información relevante para la extracción a los campos como: entidad, objeto del proceso, presupuesto referencial total, autoridades, comisión técnica, funcionario encargado del proceso, estado del proceso, tipo de adjudicación, fecha de publicación, fecha de inicio de puja, fecha posible de adjudicación. (ver imágenes 5 a 7).

Descripción	Fechas	Productos	Archivos
Descripción del Proceso de Contratación			
Entidad:	GOBIERNO AUTONOMO DESCENTRALIZADO DEL CANTON RIOBAMBA		
Objeto de Proceso :	REPUESTO PARA MAQUINARIA Y EQUIPOS DEL CAMAL MUNICIPAL		
Código:	SIE-GADMR-018-2020		
Tipo Compra:	Bien		
Presupuesto Referencial Total (Sin Iva):	USD 22,209.00		
Tipo de Contratación:	Subasta Inversa Electrónica		
Autoridades:	Cédula / Núm. Identificación	Apellidos y Nombres	Cargo
	0602037335	CADENA OLEAS BYRON NAPOLEON	Máxima Autoridad Institucional
	0601450125	HERNANDEZ CRUZ ANGEL MARCELO	Máxima Autoridad Responsable de la Gestión Administrativa - Financiera
	0602048415	JARA MEJIA DORYAN PATRICIO	Autoridad Ordenadora de Gasto
			Origen
			NACIONAL
			NACIONAL
			NACIONAL

Imagen 5: Información relevante para ser extraída.

Comisión Técnica:	Si															
Miembros Comisión Técnica	<table border="1"> <thead> <tr> <th>Cédula / Núm. Identificación</th> <th>Apellidos y Nombres</th> <th>Función en la Comisión Técnica</th> </tr> </thead> <tbody> <tr> <td>0931450274</td> <td>JACOME SEDAMANOS SHELLY SELENA</td> <td>Profesional designado por la máxima autoridad, quien lo presidirá</td> </tr> <tr> <td>1801966902</td> <td>GUERRA HURTADO HERNAN MARCELO</td> <td>Titular del área requirente</td> </tr> <tr> <td>1500655897</td> <td>TOAPANTA VILCAHUANO DIEGO GIOVANNY</td> <td>Profesional afín al objeto de la contratación designado por la máxima autoridad</td> </tr> <tr> <td>2200049464</td> <td>CHANALUISA ORELLANA MONICA YADIRA</td> <td>Secretario/a</td> </tr> </tbody> </table>	Cédula / Núm. Identificación	Apellidos y Nombres	Función en la Comisión Técnica	0931450274	JACOME SEDAMANOS SHELLY SELENA	Profesional designado por la máxima autoridad, quien lo presidirá	1801966902	GUERRA HURTADO HERNAN MARCELO	Titular del área requirente	1500655897	TOAPANTA VILCAHUANO DIEGO GIOVANNY	Profesional afín al objeto de la contratación designado por la máxima autoridad	2200049464	CHANALUISA ORELLANA MONICA YADIRA	Secretario/a
Cédula / Núm. Identificación	Apellidos y Nombres	Función en la Comisión Técnica														
0931450274	JACOME SEDAMANOS SHELLY SELENA	Profesional designado por la máxima autoridad, quien lo presidirá														
1801966902	GUERRA HURTADO HERNAN MARCELO	Titular del área requirente														
1500655897	TOAPANTA VILCAHUANO DIEGO GIOVANNY	Profesional afín al objeto de la contratación designado por la máxima autoridad														
2200049464	CHANALUISA ORELLANA MONICA YADIRA	Secretario/a														
Forma de Pago:	Anticipo: 0% Saldo: Pagos por Planilla 100.00%															
Tipo de Adjudicación:	Total															
Plazo de Entrega:	365 días															
Vigencia de Oferta:	30 días															
Funcionario encargado del proceso:	lilia.verdezoto@gporellana.gob.ec															
Estado del Proceso:	<p>Cancelado Motivo: SE CANCELA ESTE PROCESO POR VIOLACION SUSTANCIAL DE UN PROCEDIMIENTO PRECONTRACTUAL Razón: Por violación sustancial de un procedimiento precontractual Fecha de cancelación: 2021-02-25 18:06:14</p>															
Descripción:	SERVICIO DE SEGURIDAD INTEGRAL, VIGILANCIA, CUSTODIA, MONITOREO DE ALARMAS CON RESPUESTA INMEDIATA DEL SUPERVISOR PARA EL GOBIERNO AUTÓNOMO DESCENTRALIZADO DE LA PROVINCIA DE ORELLANA "GADPO"															
Costos de levantamiento de textos, reproducción y edición de los Pliegos	Costo: USD 500.00 Detalle de Pago:															
Variación mínima de la Oferta durante la Puja:	2.00% Tipo Variación: Precio total															

Imagen 6: Información relevante para ser extraída.

Descripción	Fechas	Productos	Archivos
Fechas de Control del Proceso			SIE-GADPO-CA-01-2021
Fecha de Publicación	2021-02-23 18:00:00	Indicar la fecha real en la cual desea publicar el Proceso.	
Fecha Límite de Preguntas	2021-02-26 18:00:00	Fecha máxima para solicitar aclaraciones respecto al Proceso de Contratación.	
Fecha Límite de Respuestas	2021-03-02 15:00:00	Fecha máxima para solventar cualquier inquietud relacionada al Proceso de Contratación.	
Fecha Límite entrega Ofertas	2021-03-09 15:00:00	Fecha máxima de entrega Ofertas Técnica.	
Fecha Límite solicitar Convalidación	2021-03-10 15:00:00	Fecha máxima para que la Entidad notifique los errores de forma en las ofertas.	
Fecha Límite respuesta Convalidación	2021-03-12 15:00:00	Fecha máxima para respuestas de Convalidación de Errores.	
Fecha Límite de Calificación	2021-03-15 16:00:00	Fecha máxima para calificar a proveedores.	
Fecha Inicio de Puja	2021-03-16 14:00:00	Fecha en que inicia la puja.	
Fecha Final de Puja	2021-03-16 14:15:00	Fecha en que finaliza la puja.	
Fecha Estimada de Adjudicación	2021-03-19 17:00:00	Fecha estimada para la Adjudicación de la compra.	

Imagen 7: Información relevante para ser extraída.

Continuando con la navegación web (nivel 2) se recopilará información extra de los procesos como invitaciones a proveedores, los cuales son presentados en una tabla de 20 filas, además si se encuentra disponible se ingresará a la página de resultados de negociación y resumen del contrato (ver imagen 8 a 10), en los cuales se recopilará información puntual de gran relevancia y de igual manera será almacenada para uso a futuro.

» Información Proceso Contratación

FLUJO DE PROCESO

Preguntas, Respuestas y Aclaraciones > Entrega de Propuesta > Convalidación de Errores

En Recepción

FASE PRECONTRACTUAL

- Ver Preguntas y/ó Aclaraciones
- Ver Invitaciones**
- Ver Resultados de Negociación

FASE CONTRACTUAL

- Resumen de Contrato**

Descripción | Fechas | **Productos** | Archivos

Detalle: Bienes / Servicios

Categoría Bien / Servicio

Imagen 8: Información extra del proceso

» Invitaciones a proveedores

[Regresar](#)

Invitaciones a proveedores

Código: SIE-GADPO-CA-01-2021

Objeto del proceso de contratación: SERVICIO DE SEGURIDAD INTEGRAL, VIGILANCIA, CUSTODIA, MONITOREO DE ALARMAS CON RESPUESTA INMEDIATA DEL SUPERVISOR PARA EL GOBIERNO AUTÓNOMO DESCENTRALIZADO DE LA PROVINCIA DE ORELLANA "GADPO"

Buscar Proveedor:

No.	Razón Social - Proveedor	Fecha de Invitación	Provincia - Cantón	Estado actual RUP
1	ABAD PARDO REYNALDO CESAR	2021-02-23 18:02	GUAYAS - SIMON BOLIVAR	Habilitado en RUP
2	ABARCA ABARCA JOSE RAFAEL	2021-02-23 18:02	CHIMBORAZO - RIOBAMBA	Habilitado en RUP
3	ABARCA SEGURIDAD ABARCOMSEG CIA.LTDA.	2021-02-23 18:02	CHIMBORAZO - RIOBAMBA	Habilitado en RUP
4	ACHAN AQUINO JOSE IGNACIO	2021-02-23 18:02	AZUAY - CUENCA	No Habilitado en RUP
5	ACKER S.A.	2021-02-23 18:02	GUAYAS - GUAYAQUIL	Habilitado en RUP
6	ACUÑA & DELGADO ACUÑAYDELGADO CIA.LTDA.	2021-02-23 18:02	GUAYAS - DURAN	Habilitado en RUP
7	AEROSEG SERVICIO TECNICO DE SEGURIDAD PRIVADA CIA. LTDA	2021-02-23 18:02	PICHINCHA - QUITO	Habilitado en RUP
8	A.E.S.M. ASESORES ESPECIALISTAS EN SEGURIDAD Y MONITOREO CIA. LTDA.	2021-02-23 18:02	PICHINCHA - QUITO	Habilitado en RUP
9	AGENCIA DE PROTECTORES PRIVADOS DEL ECUADOR MORAN AGEPROECA CIA. LTDA.	2021-02-23 18:02	GUAYAS - GUAYAQUIL	No Habilitado en RUP
10	AGENCIA DE SEGURIDAD PRIVADA ASEGPRIV CIA.LTDA.	2021-02-23 18:02	AZUAY - CUENCA	Habilitado en RUP
11	AGENCIA DE SEGURIDAD Y VIGILANCIA PRIVADA ASEVIP CIA.LTDA.	2021-02-23 18:02	GUAYAS - GUAYAQUIL	Habilitado en RUP
12	AGENTES DE SEGURIDAD Y VIGILANCIA AGENTESEG CIA.LTDA.	2021-02-23 18:02	IMBABURA - IBARRA	Habilitado en RUP
13	AGUICENTINELA SEGURIDAD PRIVADA COMPAÑIA LIMITADA	2021-02-23 18:02	PICHINCHA - QUITO	Habilitado en RUP
14	AGUILAR ROMERO WILLAN GERARDO	2021-02-23 18:02	SUCUMBIOS - LAGO AGRIO	Habilitado en RUP
15	AGUILASDORADASSESECURITY CIA. LTDA.	2021-02-23 18:02	CARCHI - MONTUFAR	No Habilitado en RUP
16	AGUISEP CIA.LTDA.	2021-02-23 18:02	AZUAY - CUENCA	Habilitado en RUP
17	AGUSEGPRO CIA. LTDA. AGUILAS, SEGURIDAD & PROTECCION	2021-02-23 18:02	GUAYAS - GUAYAQUIL	Habilitado en RUP
18	AJON AGUINDA ZANDRA LIDIA	2021-02-23 18:02	ORELLANA - FRANCISCO DE ORELLANA	Habilitado en RUP
19	ALARCON MORENO BOLIVAR RAFAEL	2021-02-23 18:02	GUAYAS - SAMBORONDON	Habilitado en RUP
20	ALARMAS CONTROL Y SEGURIDAD ALCONSEG CIA. LTDA.	2021-02-23 18:02	GUAYAS - GUAYAQUIL	Habilitado en RUP

[Siguiente](#) [Fin](#)

Proveedores del 1 al 20 de 1192

Imagen 9: Lista de proveedores invitados a participar

Información del Producto Adjudicado al Proveedor						
Producto (CPC)	Descripción	Tiempo de Entrega (días)	Cantidad	Precio Unitario (\$)	% Adjudicado	Total Adjudicado (\$)
44253.00.1	PARTES Y PIEZAS DE LAS HERRAMIENTAS DE USO MANUAL, NEUMATICAS O CON MOTOR NO ELECTRICO INCORPORADO: SIERRAS O TROZADORAS DE CADENA, ETC.	60	1.00000 (Global)	20,654.37	100.00	20,654.37

Información de Adjudicación						
#	RUC	Nombre del Adjudicatario	Tipo	Fecha de Adjudicación	Monto de Adjudicación (\$)	Tipo de Adjudicación
1	0601399579001	OROZCO ARIAS LUIS FERNANDO	NATURAL	2020-09-28	20,654.37	TOTAL

Imagen 10: Información relevante del resumen de contrato.

A partir del nivel 1 no se requiere la solución de un sistema de CAPTCHA por ende la tarea del Web Crawler se ve facilitada y con posibilidad de navegar libremente entre los niveles de información de cada proceso. En la ilustración 2 se simplifica el proceso cíclico a seguir por el Web Crawler para obtener cada proceso. En la ilustración 3 se presenta el flowchart del funcionamiento del Web Crawler.

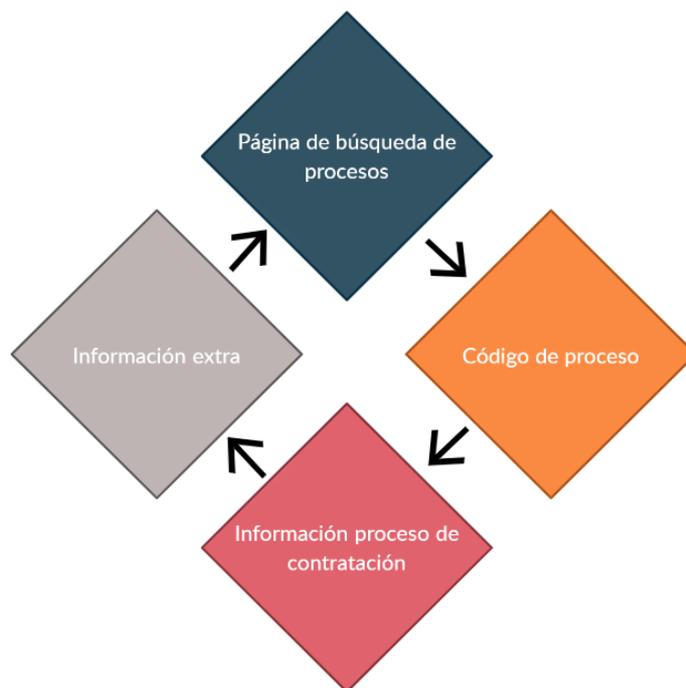


Ilustración 2: Proceso cíclico de obtención de procesos.

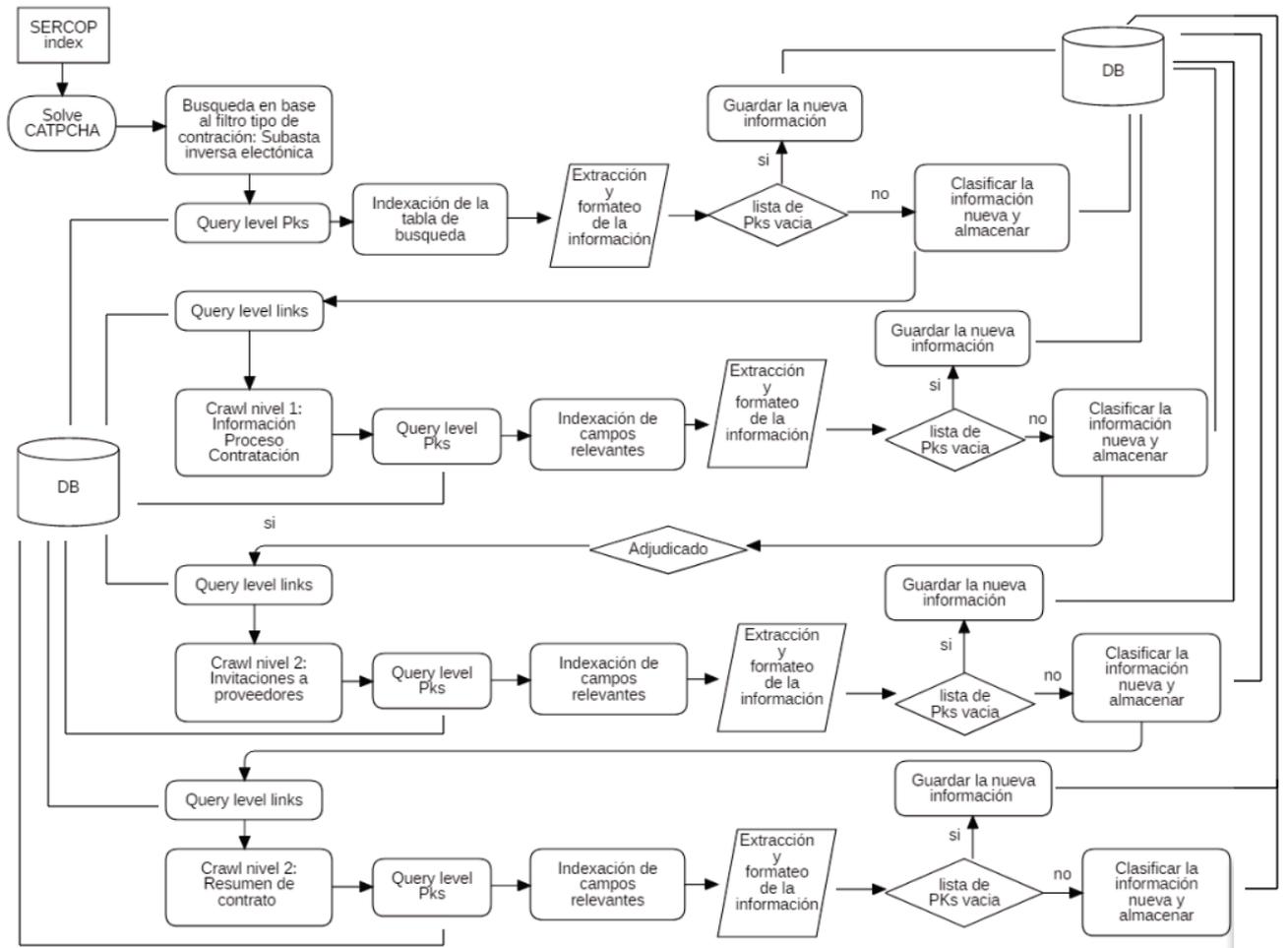


Ilustración 3: Flowchart del Web Crawler

Web Crawler

Se entiende como Web Crawler a un Bot de internet cuya funcionalidad principal es el escaneo automático de la información contenida en una web para posteriormente obtener los datos mediante técnicas de scraping y finalmente crear un índice de los datos obtenidos, generalmente mediante la implementación de bases de datos (octoparse, 2019). En el presente proyecto se elaboró un Web Crawler específico para el portal de compras públicas, el desarrollo fue en base al lenguaje de programación Python v3.6 y el módulo de automatización web Selenium.

Para el funcionamiento del Web Crawler es necesario establecer los filtros correctos y solucionar el CAPTCHA inicial, ya que en el presente proyecto no se aborda la solución automatizada de este tipo de filtros de seguridad.

Dentro de las funcionalidades del Web Crawler implementado se encuentran: la navegación automatizada en las páginas web de la SERCOP, extracción y almacenamiento de información y documentos relevantes de cada proceso de contratación mediante subasta inversa electrónica.

Con el objetivo de optimizar el funcionamiento del Web Crawler implementado, se procedió a dividir el trabajo en 3 módulos principales cada uno encargado de una tarea específica. El primer módulo será el de navegación, el cual se encargará de realizar el crawling¹ por los distintos niveles del portal web de la SERCOP y localizará la información de relevancia para ser extraída. El segundo módulo será el encargado de la extracción de información relevante previamente señalada por el primero módulo y mediante técnicas de scraping² se encargará de estructurar todos los datos recuperados. Finalmente, el tercer y último módulo

¹ Hace referencia a cualquier Bot de indexación enviado por los motores de búsqueda, con el fin de detectar, leer y analizar todo el contenido y el código que compone a una página web realizando los correspondientes saltos a través de los enlaces de cada web (Neoattack, s/f).

² El web scraping (del inglés scraping = arañar/raspar) se extraen y almacenan datos de páginas web para analizarlos o utilizarlos en otra parte. Por medio de este raspado web se almacenan diversos tipos de información (Ionos, 2020).

será el encargado de almacenar toda la información recopilada por el segundo módulo en las bases de datos correspondientes para su futuro uso. Toda la documentación del programa se encuentra en el anexo A incluido en el documento. En cuanto al desarrollo del software se tiene en cuenta el caso en que el formato de la página web de la cual se desea extraer la información cambie, es por eso que se trató de hacer lo más flexible posible al funcionamiento del WECPSIE con el objetivo de poder parchar rápidamente a futuro en caso de ser necesario. En la ilustración 3 se detalla el funcionamiento de cada módulo implementado en el Web Crawler.

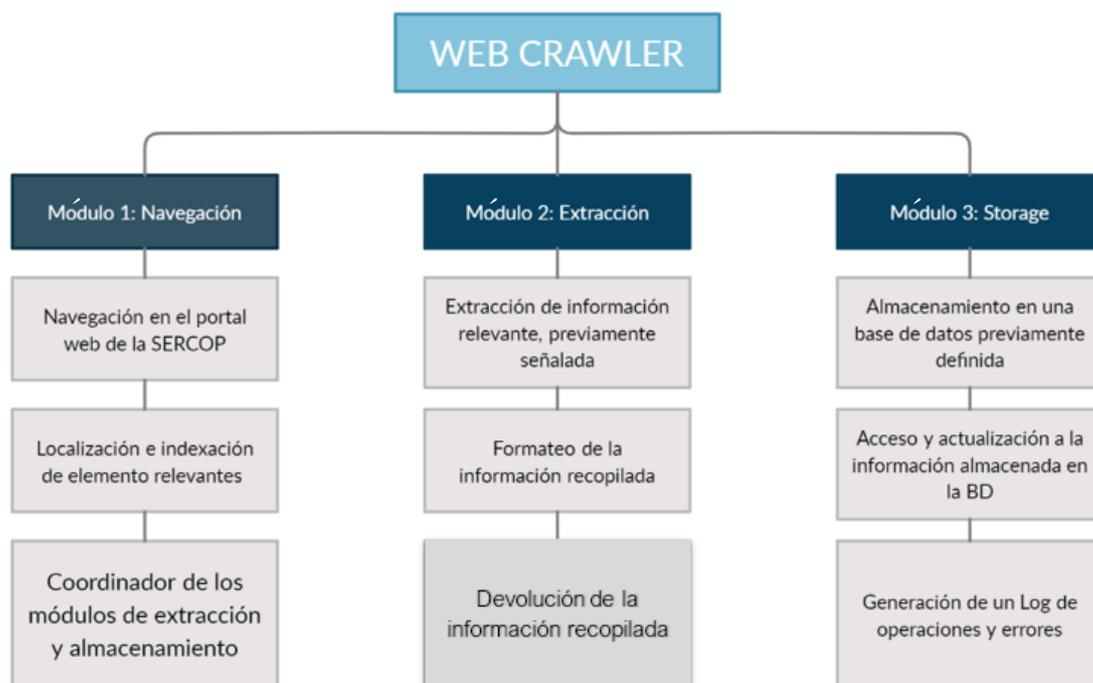


Ilustración 4: Módulos del Web Crawler

Módulo de navegación

Este módulo es el encargado de la apertura y navegación en las diferentes páginas web correspondientes a cada proceso de SIE, en cada ventana se encargará de la búsqueda de elementos claves los cuales a su vez serán dirigidos al módulo de extracción y posteriormente

al módulo de almacenamiento. El presente módulo se encuentra compuesto de cuatro funciones cada una de ellas correspondiente al nivel de navegación en el que se encuentre el software.

Nivel	Método	Página
Index	swing_index_pages	Búsqueda de procesos SIE
1	crawl_pcpsie_page	Información Proceso Contratación
2	crawl_pcpsie_contrato_page	Resumen Información Esencial del Contrato
2	crawl_pcpsie_invitacion_page	Invitaciones a proveedores

Tabla 1: Funciones del módulo de navegación

El software WECPSIE comienza con el nivel denominado Index en el cual se encuentra los parámetros de búsqueda para los diferentes tipos de procesos de contratación, en este caso el enfoque es en los procesos SIE. Una vez abierta la página web correspondiente, el software selecciona de manera automática el tipo *Subasta Inversa Electrónica* en la casilla *Tipo de contratación* y además establece el rango de fechas dentro del cual se obtendrán los procesos. Cabe mencionar, que el límite máximo a partir de una fecha final es de 6 meses atrás en el tiempo, es decir si se selecciona 1 de enero de 2021 como fecha final solo se lograrán obtener los procesos hasta el 1 de julio de 2020 (ver imagen 11). Una vez establecidos los parámetros se otorga al usuario un periodo de 12 segundos en los cuales se deberá completar correctamente el CAPTCHA, caso contrario el programa finalizará.

Una vez completado de manera correcta el CAPTCHA el software ejecutará el botón de búsqueda, dando como resultado el despliegue de una tabla de al menos 20 filas en la parte inferior de la pantalla, la cual contendrá la información principal de cada proceso; además, se localizará los índices de dicha tabla lo cual será útil para indicar al software cuando finalizar la primera etapa. Una vez localizada la tabla de procesos presentada, el módulo de navegación procede a enviar la misma al módulo de extracción; concretamente a la función

get_main_page_table_info la cual será la encargada de extraer el valor de cada columna y devolver en una representación estructurada todos los datos de cada uno de los registros.

Una vez terminada la función del módulo de extracción de inmediato el módulo de navegación se encarga de pasar los datos devueltos al módulo de almacenamiento, previamente comprobando si en la base de datos ya existen registros de los procesos obtenidos o no. Con ello se determina el Query correcto a ejecutar; Update ó Insert.

Código	Entidad Contratante	Objeto del Proceso	Estado del Proceso	Provincial/Cantón	Presupuesto Referencial Total(sin iva)	Fecha de Publicación	Opciones
CDC-GADPRPN-004-2020	JUNTA PARROQUIAL PUERTO NAPO	CONSULTORÍA PARA LA ELABORACIÓN DE LOS ESTUDIOS Y DISEÑOS DEFINITIVOS PARA LA REGENERACIÓN URBANA DE LAS CALLES DE LA CABECERA PARROQUIAL DE LA PARROQUIA PUERTO NAPO.	Desierta	NAPO / TENA	\$33,035.52	2020-10-12 14:00:00	
CDC-MNE-006-2020	MANCOMUNIDAD DEL NORTE DEL ECUADOR	Formular el plan estratégico integral y Modelo de gestión de la "Mancomunidad de los Gobiernos Autónomos Descentralizados Provinciales del Norte del Ecuador", MNE., que permita generar un desarrollo fronterizo integral tomando en cuenta el territorio, los actores y el marco jurídico e institucional que regula su gestión.	Desierta	CARCHI / TULCAN	\$13,392.86	2020-10-12 13:00:00	
LICBS-CELTGM-049-20	CORPORACION ELECTRICA DEL ECUADOR CELEC EP.	TGM: ADQUISICIÓN DE PARTES DE RUTA DE GASES CALIENTES DE LAS UNIDADES M6001FA	Desierta	AZUAY / CUENCA	\$1,567,992.60	2020-10-12 14:00:00	
LICO-EPMBPCD-2020-02	Empresa Pública Municipal Registro de la Propiedad del Cantón Durán	CONSTRUCCIÓN DEL EDIFICIO EMPRESA PÚBLICA MUNICIPAL REGISTRO DE LA PROPIEDAD DEL CANTÓN DURÁN	Ejecución de Contrato	GUAYAS / DURAN	\$3,194,668.91	2020-10-12 13:00:00	
MCBS-SRIDZZ-01-2020	SERVICIO DE RENTAS INTERNAS	SERVICIO DE TAXI PARA LA MOVILIZACIÓN DE LOS SERVIDORES Y FUNCIONARIOS DE LA DIRECCIÓN ZONAL 2 DEL SERVICIO DE RENTAS INTERNAS, PROVINCIA DE NAPO	Ejecución de Contrato	PICHINCHA / QUITO	\$16,200.00	2020-10-12 13:00:00	

Imagen 11: Información a extraer en el nivel Index

Finalizada la extracción del nivel Index el módulo de navegación obtiene de la base de datos los enlaces correspondientes a la información de cada proceso y procede a abrir cada uno de ellos de manera iterativa. Esta ventana corresponde al nivel 1 de navegación (ver imagen 12), en la cual se presenta dentro de una tabla toda la información correspondiente al proceso. Dependiendo del estado en el que se encuentre el mismo, será factible encontrar un enlace a la información acerca del contrato de adjudicación y un enlace a los proveedores que fueron invitados a participar en el proceso. De igual manera una vez localizada la tabla correspondiente el crawler la envía al módulo de extracción más precisamente a la función *extract_SIE_lv1*, la misma será la encargada de extraer la información del proceso que fue

considerada de mayor relevancia y al finalizar devolverá correctamente estructurada y organizada toda la información recopilada. Finalmente, el módulo de navegación se encargará de enviar al módulo de almacenamiento la información obtenida de cada proceso, previamente comprobando si en la base de datos ya existen los datos correspondientes al registro o no y en cada caso se ejecutará un Query de Update o Insert correspondientemente.

Descripción del Proceso de Contratación																	
Entidad:	HOSPITAL GENERAL EMERALDAS SUR-DELFINA TORRES DE CONCHA																
Objeto de Proceso :	ADQUISICIÓN DE DISPOSITIVOS MÉDICOS DE LABORATORIOS PARA EL CENTRO DE HEMODIALISIS DEL HOSPITAL GENERAL EMERALDAS SUR DELFINA TORRES DE CONCHA																
Código:	SIE-HCESDTC-036-2020																
Tipo Compra:	Bien																
Presupuesto Referencial Total (Sin Iva):	USD 151,971.96																
Tipo de Contratación:	Subasta Inversa Electrónica																
Autoridades:	<table border="1"> <thead> <tr> <th>Cédula / Núm. Identificación</th> <th>Apellidos y Nombres</th> <th>Cargo</th> <th>Origen</th> </tr> </thead> <tbody> <tr> <td>1709561508</td> <td>CEDEÑO SOLIS ROBERTH NIMIO</td> <td>Máxima Autoridad Institucional</td> <td>NACIONAL</td> </tr> <tr> <td>0802352674</td> <td>GOMEZ ESTUPIÑAN JUAN PABLO</td> <td>Máxima Autoridad Responsable de la Gestión Administrativa - Financiera</td> <td>NACIONAL</td> </tr> <tr> <td>1709561508</td> <td>CEDEÑO SOLIS ROBERTH NIMIO</td> <td>Autoridad Ordenadora de Gasto</td> <td>NACIONAL</td> </tr> </tbody> </table>	Cédula / Núm. Identificación	Apellidos y Nombres	Cargo	Origen	1709561508	CEDEÑO SOLIS ROBERTH NIMIO	Máxima Autoridad Institucional	NACIONAL	0802352674	GOMEZ ESTUPIÑAN JUAN PABLO	Máxima Autoridad Responsable de la Gestión Administrativa - Financiera	NACIONAL	1709561508	CEDEÑO SOLIS ROBERTH NIMIO	Autoridad Ordenadora de Gasto	NACIONAL
Cédula / Núm. Identificación	Apellidos y Nombres	Cargo	Origen														
1709561508	CEDEÑO SOLIS ROBERTH NIMIO	Máxima Autoridad Institucional	NACIONAL														
0802352674	GOMEZ ESTUPIÑAN JUAN PABLO	Máxima Autoridad Responsable de la Gestión Administrativa - Financiera	NACIONAL														
1709561508	CEDEÑO SOLIS ROBERTH NIMIO	Autoridad Ordenadora de Gasto	NACIONAL														
Comisión Técnica:	Si																
Miembros Comisión Técnica	<table border="1"> <thead> <tr> <th>Cédula / Núm. Identificación</th> <th>Apellidos y Nombres</th> <th>Función en la Comisión Técnica</th> </tr> </thead> <tbody> <tr> <td>1803023827</td> <td>ARBOLEDA MAZON ALEX MAURICIO</td> <td>Profesional designado por la máxima autoridad, quien lo presidirá</td> </tr> <tr> <td>0803036672</td> <td>PAZ MORAN VERONICA ALEXANDRA</td> <td>Titular del área requirente</td> </tr> <tr> <td>0803231893</td> <td>GRACIA CHAMORRO RODOLFO RUBEN</td> <td>Profesional afín al objeto de la contratación designado por la máxima autoridad</td> </tr> <tr> <td>0803102342</td> <td>BEDON MARIN DORIS MERCEDES</td> <td>Secretario/a</td> </tr> </tbody> </table>	Cédula / Núm. Identificación	Apellidos y Nombres	Función en la Comisión Técnica	1803023827	ARBOLEDA MAZON ALEX MAURICIO	Profesional designado por la máxima autoridad, quien lo presidirá	0803036672	PAZ MORAN VERONICA ALEXANDRA	Titular del área requirente	0803231893	GRACIA CHAMORRO RODOLFO RUBEN	Profesional afín al objeto de la contratación designado por la máxima autoridad	0803102342	BEDON MARIN DORIS MERCEDES	Secretario/a	
Cédula / Núm. Identificación	Apellidos y Nombres	Función en la Comisión Técnica															
1803023827	ARBOLEDA MAZON ALEX MAURICIO	Profesional designado por la máxima autoridad, quien lo presidirá															
0803036672	PAZ MORAN VERONICA ALEXANDRA	Titular del área requirente															
0803231893	GRACIA CHAMORRO RODOLFO RUBEN	Profesional afín al objeto de la contratación designado por la máxima autoridad															
0803102342	BEDON MARIN DORIS MERCEDES	Secretario/a															
Forma de Pago:	Anticipo: 0% Saldo: Pago contra entrega de bienes obras o servicio 100.00%																
Tipo de Adjudicación:	Total																
Plazo de Entrega:	15 días																
Vigencia de Oferta:	60 días																
Funcionario encargado del proceso:	doris.bedon@hdtc.gob.ec																
Estado del Proceso:	Adjudicado - Registro de Contratos																
Descripción:	ADQUISICIÓN DE DISPOSITIVOS MÉDICOS DE LABORATORIOS PARA EL CENTRO DE HEMODIALISIS DEL HOSPITAL GENERAL EMERALDAS SUR DELFINA TORRES DE CONCHA																
Variación mínima de la Oferta durante la Puja:	1.00% Tipo Variación: Precio total																

Imagen 12: Página correspondiente al nivel 1 de navegación

El nivel 2 de navegación corresponde a las páginas de información del contrato de adjudicación en caso de existir y la página de invitaciones a proveedores. El módulo de navegación será el encargado de consultar a la base de datos los enlaces correspondientes a cada página y en caso de existir los procederá a abrir en una ventana del navegador.

Dentro de la página Invitaciones a proveedores se presenta una tabla con la información de cada proveedor muy similar a la tabla del nivel Index en la cual el número de filas puede variar entre cero y miles. El crawler enviará a la función *extract_SIE_lv2* dentro del módulo de almacenamiento la tabla con los proveedores y al finalizar la extracción de los datos de cada fila la función devuelve correctamente estructurada la información obtenida (ver imagen 13).

Posterior a ello el módulo de navegación comprueba en la base de datos si existen registros previos correspondientes al proceso y almacena la información, con la ayuda del módulo de almacenamiento, mediante un Query de Update o Insert según sea el caso.

Por otro lado, dependiendo del estado en el cual se encuentre el proceso de contratación puede existir un enlace a la página con la información acerca del contrato de adjudicación (ver imagen 14). Por lo cual el módulo de navegación consulta todos los enlaces disponibles en la base de datos y los procede a abrir. En la página de información del contrato se presentan varias tablas sin embargo en este caso se toma los datos contenidos dentro de la tabla información de adjudicación ya que se los considera muy relevantes. La función *extract_SIE_lv2_inv* ubicada en el módulo de extracción es la encargada de obtener y devolver de manera estructurada todos los datos presentes dentro de la tabla.

De igual manera el módulo de navegación comprueba en la base de datos si existen registros previos correspondientes al proceso en turno o no y el módulo de almacenamiento procede a guardar la información mediante un Query de Update o Insert según sea el caso.

Invitaciones a proveedores				
Código:		SIE-CNEZCH-002-2021		
Objeto del proceso de contratación:		CONTRATACIÓN DEL SERVICIO DE SEGURIDAD Y VIGILANCIA PARA LA DELEGACIÓN PROVINCIAL ELECTORAL DE ZAMORA CHINCHIPE		
Buscar Proveedor: <input checked="" type="radio"/> Por RUC <input type="radio"/> Por Razón Social		<input type="text"/> <input type="button" value="Buscar Proveedor"/>		
No.	Razón Social - Proveedor	Fecha de Invitación	Provincia - Cantón	Estado actual RUP
1	ABAD PARADO REYNALDO CESAR	2021-04-08 20:02	GUAYAS - SIMÓN BOLÍVAR	Habilitado en RUP
2	ABARCA ABARCA JOSE RAFAEL	2021-04-08 20:02	CHIMBORAZO - RIOBAMBA	Habilitado en RUP
3	ABARCA SEGURIDAD ABARCOMSEG CIA.LTDA.	2021-04-08 20:02	CHIMBORAZO - RIOBAMBA	Habilitado en RUP
4	ACHAN AQUINO JOSE IGNACIO	2021-04-08 20:02	AZUAY - CUENCA	No Habilitado en RUP
5	ACKER S.A.	2021-04-08 20:02	GUAYAS - GUAYAQUIL	Habilitado en RUP
6	ACUÑA & DELGADO ACUÑAYDELGADO CIA.LTDA.	2021-04-08 20:02	GUAYAS - DURAN	No Habilitado en RUP
7	AEROSSEG SERVICIO TECNICO DE SEGURIDAD PRIVADA CIA. LTDA	2021-04-08 20:02	PICHINCHA - QUITO	Habilitado en RUP
8	A.E.S.M. ASESORES ESPECIALISTAS EN SEGURIDAD Y MONITOREO CIA. LTDA.	2021-04-08 20:02	PICHINCHA - QUITO	Habilitado en RUP
9	AGENCIA DE PROTECTORES PRIVADOS DEL ECUADOR MORAN AGEPROECUA CIA. LTDA.	2021-04-08 20:02	GUAYAS - GUAYAQUIL	No Habilitado en RUP
10	AGENCIA DE SEGURIDAD PRIVADA ASESPRV CIA.LTDA.	2021-04-08 20:02	AZUAY - CUENCA	Habilitado en RUP
11	AGENCIA DE SEGURIDAD Y VIGILANCIA PRIVADA ASEVIP CIA.LTDA.	2021-04-08 20:02	GUAYAS - GUAYAQUIL	Habilitado en RUP
12	AGENTES DE SEGURIDAD Y VIGILANCIA AGENTESEG CIA.LTDA.	2021-04-08 20:02	IMBABURA - IBARRA	Habilitado en RUP
13	AGUICENTINELA SEGURIDAD PRIVADA COMPAÑIA LIMITADA	2021-04-08 20:02	PICHINCHA - QUITO	No Habilitado en RUP
14	AGUILAR ROMERO WILLAN GERARDO	2021-04-08 20:02	SUCUMBIOS - LAGO AGRIO	Habilitado en RUP
15	AGUILASORADASSESECURITY CIA. LTDA.	2021-04-08 20:02	CARCHI - MONTUFAR	No Habilitado en RUP
16	AGUISEP CIA.LTDA.	2021-04-08 20:02	AZUAY - CUENCA	Habilitado en RUP
17	AGUISEPRO CIA. LTDA. AGUILAS, SEGURIDAD & PROTECCION	2021-04-08 20:02	GUAYAS - GUAYAQUIL	Habilitado en RUP
18	AJON AGUINDA ZANDRA LIDIA	2021-04-08 20:02	ORELLANA - FRANCISCO DE ORELLANA	Habilitado en RUP
19	ALARCON MORENO BOLIVAR RAFAEL	2021-04-08 20:02	GUAYAS - SAMBORONDON	Habilitado en RUP
20	ALARINAS CONTROL Y SEGURIDAD ALCONSEG CIA. LTDA.	2021-04-08 20:02	GUAYAS - GUAYAQUIL	Habilitado en RUP

Imagen 13: Página de invitaciones a proveedores

Resumen Información Esencial del Contrato

Regresar

Información Esencial del Proceso de Contratación

Entidad Contratante: INIAP ESTACION EXPERIMENTAL SANTA CATALINA
 Código del Proceso: SIE-INIAPSC-013-2020
 Objeto de Contratación: ADQUISICIÓN E INSTALACIÓN DE UN "SISTEMA ÓPTICO PARA DETECTOR DE FLUORESCENCIA (FLO) DEL EQUIPO DE CROMATOGRAFÍA LÍQUIDA DE ALTA RESOLUCIÓN (HPLC)" DEL LABORATORIO LASA DEL DEPARTAMENTO NUTRICIÓN Y CALIDAD DE LA ESTACIÓN EXPERIMENTAL SANTA CATALINA
 Tipo de Compra: Bida
 Tipo de Procedimiento: Subasta Inversa Electrónica
 Presupuesto Referencial: \$ 14.887,06
 Anticipo: Porcentaje: 0,00 %
 Valor: \$ 0,00
 Saldo: Porcentaje: 100,00 %
 Valor: \$ 14.887,06
 Plazo de entrega: 30 días

Información de Adjudicación

#	RUC	Nombre del Adjudicatario	Fecha de Adjudicación	Monto de Adjudicación (\$)
1	1791316290001	SOLUCIONES TÉCNICAS Y CORPORATIVAS FIELQUÍMEC CIA. LTDA.	2020-10-29	13.826,37

Información del Contrato

#	RUC	Razón Social	Monto Contractual (\$)	Plazo Contractual (días)	Rego Contra Entrega	Renuncia de Anticipo	Anticipo		El Contrato inicia en	Fecha de Suscripción del Contrato	Fecha de notificación de disponibilidad de anticipo	Fecha de otra condición de anulado a la naturaleza del contrato	Número de entrega parcial de contrato	Fecha de entrega de información de contrato	Administración del Contrato (Usuario / Nombre)
							Porcentaje (%)	Valor (\$)							
1	1791316290001	SOLUCIONES TÉCNICAS Y CORPORATIVAS FIELQUÍMEC CIA. LTDA.	13.826,37	30	SI	NO	0,00	0,00	Fecha de Suscripción de Contrato	2020-11-05	-	-	Entrega Definitiva	2020-12-05	tsamanagon - IVAN RODRIGO SAMAÑOSO MALQUA

Garantías

#	RUC	Razón Social	Nro.	Garantía	Forma	Otorgante	Monto Garantía (\$)	Número Documento	Fecha Otorgamiento	Vigencia (días)	Fecha Vencimiento	Descripción
1	1791316290001	SOLUCIONES TÉCNICAS Y CORPORATIVAS FIELQUÍMEC CIA. LTDA.	1	Técnica	Documento	FIELQUÍMEC	-	001	2020-11-19	360	2021-11-19	OT
			2	Técnica	Documento	FIELQUÍMEC	-	001	-	360	2022-04-04	-

Entregas Parciales

#	RUC	Razón Social	Nro.	Nombre	Fecha
1	1791316290001	SOLUCIONES TÉCNICAS Y CORPORATIVAS FIELQUÍMEC CIA. LTDA.	1	Entrega Definitiva	2020-12-05

Documentación Relevante y Obligatoria

#	RUC	Razón Social	Nro.	Documento	Descripción	Nombre Original	Fecha
1	1791316290001	SOLUCIONES TÉCNICAS Y CORPORATIVAS FIELQUÍMEC CIA. LTDA.	1	Compro. Suscripción de la JAMES	COMPROBANTE	COMPROBANTE SUSCRIPCION.pdf	2020-11-05 11:48:47
			2	Documento de Garantía	GARANTIA TECNICA	GARANTIA TECNICA.pdf	2020-11-05 13:24:15
			3	Documento de Garantía	GARANTIA TECNICA	GARANTIA TECNICA.pdf	2020-12-10 08:01:18

Imagen 14: Página de información esencial del contrato

Módulo de extracción

A partir de los elementos web entregados por parte del módulo de navegación; el módulo de extracción se encargará de obtener la información señalada como relevante. Dentro del módulo existen cuatro funciones (ver tabla 2) las cuales serán invocadas en cada nivel en específico, al finalizar la extracción se devolverá de manera estructurada los datos recopilados mediante diccionarios para que la manipulación de los datos en el módulo de almacenamiento sea mucho más sencilla. Es importante mencionar que para la extracción de los niveles 1 y 2 se configuró un parámetro para que el programa realice la extracción sin la necesidad de la apertura de una ventana del navegador Chrome.

Nivel	Método	Objeto Web	Información Extraída
Index	Get_main_page_table_info	Tabla de búsqueda	Texto y enlaces
1	Extract_SIE_lv1	Tabla descripción del proceso y tabla de fase contractual	Texto y enlaces
2	Extract_SIE_lv2	Tabla información de contrato de adjudicación	Texto
2	Extract_SIE_lv2_inv	Tabla de invitaciones a proveedores	Texto

Tabla 2: Funciones en el módulo de extracción

Dependiendo el objeto web del cual se necesitó extraer la información se utilizaron varios métodos de localización y extracción. En caso de ser una tabla se iteró por todas sus filas y columnas extrayendo toda la información necesaria (ver imagen 15 a 19). Por otro lado, la mayoría de los elementos web en el portal de compras públicas no cuentan con un ID único por lo cual se dificultaba la extracción de los datos en ciertas ocasiones, es por ello que se utilizó los Xpath de los elementos directamente para proceder a la extracción del campo.

Código	Entidad Contratante	Objeto del Proceso	Estado del Proceso	Provincia/Cantón	Presupuesto Referencial Total (sin Iva)	Fecha de Publicación	Opciones
CDC-GADPRPI-004-2020	JUNTA PARROQUIAL PUERTO NAPO	CONSULTORÍA PARA LA ELABORACIÓN DE LOS ESTUDIOS Y DISEÑOS DEFINITIVOS PARA LA REGENERACION URBANA DE LAS CALLES DE LA CABECERA PARROQUIAL DE LA PARROQUIA PUERTO NAPO.	Desierta	NAPO / TENA	\$33,035.52	2020-10-12 14:00:00	
CDC-MNE-006-2020	MANCOMUNIDAD DEL NORTE DEL ECUADOR	Formular el plan estratégico integral y Modelo de gestión de la "Mancomunidad de los Gobiernos Autónomos Descentralizados Provinciales del Norte del Ecuador", MNE., que permita generar un desarrollo fronterizo integral tomando en cuenta el territorio, los actores y el marco jurídico e institucional que regula su gestión.	Desierta	CARCHI / TULCAN	\$13,392.86	2020-10-12 13:00:00	
LICBS-CELTGM-049-20	CORPORACION ELECTRICA DEL ECUADOR CELEC EP	TGM: ADQUISICIÓN DE PARTES DE RUTA DE GASES CALIENTES DE LAS UNIDADES MS6001FA	Desierta	AZUAY / CUENCA	\$1,567,992.60	2020-10-12 14:00:00	
LICO-EPMRPCD-2020-02	Empresa Pública Municipal Registro de la Propiedad del Cantón Durán	CONSTRUCCIÓN DEL EDIFICIO EMPRESA PÚBLICA MUNICIPAL REGISTRO DE LA PROPIEDAD DEL CANTÓN DURÁN	Ejecución de Contrato	GUAYAS / DURAN	\$3,194,668.91	2020-10-12 13:00:00	
MCBS-SRIDZ2-01-2020	SERVICIO DE RENTAS INTERNAS	SERVICIO DE TAXI PARA LA MOVILIZACIÓN DE LOS SERVIDORES Y FUNCIONARIOS DE LA DIRECCIÓN ZONAL 2 DEL SERVICIO DE RENTAS INTERNAS, PROVINCIA DE NAPO	Ejecución de Contrato	PICHINCHA / QUITO	\$16,200.00	2020-10-12 13:00:00	
MCO-GMJ-DOP-19-2020	GOBIERNO AUTONOMO DESCENTRALIZADO MUNICIPAL DEL CANTON JIPIJAPA	CONSTRUCCIÓN DE DUCTO CAJÓN DE HORMIGÓN ARMADO F ^o C=240 KG EN LA CIUDADELA DE LOS MAESTROS DE LA PARROQUIA PARRALES Y GUALE DEL CANTÓN JIPIJAPA	Ejecución de Contrato	MANABI / JIPIJAPA	\$44,642.86	2020-10-12 14:00:00	

Imagen 15: Información extraída del nivel Index

Descripción del Proceso de Contratación																	
Entidad:	HOSPITAL GENERAL ESMERALDAS SUR-DELFINA TORRES DE CONCHA																
Objeto de Proceso :	ADQUISICIÓN DE DISPOSITIVOS MÉDICOS DE LABORATORIOS PARA EL CENTRO DE HEMODIALISIS DEL HOSPITAL GENERAL ESMERALDAS SUR DELFINA TORRES DE CONCHA																
Código:	SIE-HGESDTC-036-2020																
Tipo Compra:	Bien																
Presupuesto Referencial Total (Sin Iva):	USD 151,971.96																
Tipo de Contratación:	Subasta Inversa Electrónica																
Autoridades:	<table border="1"> <thead> <tr> <th>Cédula / Núm. Identificación</th> <th>Apellidos y Nombres</th> <th>Cargo</th> <th>Origen</th> </tr> </thead> <tbody> <tr> <td>1709561508</td> <td>CEDEÑO SOLIS ROBERTH NIMIO</td> <td>Máxima Autoridad Institucional</td> <td>NACIONAL</td> </tr> <tr> <td>0802352674</td> <td>COMEZ ESTUPIÑAN JUAN PABLO</td> <td>Máxima Autoridad Responsable de la Gestión Administrativa - Financiera</td> <td>NACIONAL</td> </tr> <tr> <td>1709561508</td> <td>CEDEÑO SOLIS ROBERTH NIMIO</td> <td>Autoridad Ordenadora de Gasto</td> <td>NACIONAL</td> </tr> </tbody> </table>	Cédula / Núm. Identificación	Apellidos y Nombres	Cargo	Origen	1709561508	CEDEÑO SOLIS ROBERTH NIMIO	Máxima Autoridad Institucional	NACIONAL	0802352674	COMEZ ESTUPIÑAN JUAN PABLO	Máxima Autoridad Responsable de la Gestión Administrativa - Financiera	NACIONAL	1709561508	CEDEÑO SOLIS ROBERTH NIMIO	Autoridad Ordenadora de Gasto	NACIONAL
Cédula / Núm. Identificación	Apellidos y Nombres	Cargo	Origen														
1709561508	CEDEÑO SOLIS ROBERTH NIMIO	Máxima Autoridad Institucional	NACIONAL														
0802352674	COMEZ ESTUPIÑAN JUAN PABLO	Máxima Autoridad Responsable de la Gestión Administrativa - Financiera	NACIONAL														
1709561508	CEDEÑO SOLIS ROBERTH NIMIO	Autoridad Ordenadora de Gasto	NACIONAL														
Comisión Técnica:	Si																
Miembros Comisión Técnica	<table border="1"> <thead> <tr> <th>Cédula / Núm. Identificación</th> <th>Apellidos y Nombres</th> <th>Función en la Comisión Técnica</th> </tr> </thead> <tbody> <tr> <td>1803023827</td> <td>ARBOLEDA MAZON ALEX MAURICIO</td> <td>Profesional designado por la máxima autoridad, quien lo presidirá</td> </tr> <tr> <td>0803036672</td> <td>PAZ MORAN VERONICA ALEXANDRA</td> <td>Titular del área requirente</td> </tr> <tr> <td>0803231893</td> <td>GRACIA CHAMORRO RODOLFO RUBEN</td> <td>Profesional afín al objeto de la contratación designado por la máxima autoridad</td> </tr> <tr> <td>0803102342</td> <td>BEDON MARIN DORIS MERCEDES</td> <td>Secretario/a</td> </tr> </tbody> </table>	Cédula / Núm. Identificación	Apellidos y Nombres	Función en la Comisión Técnica	1803023827	ARBOLEDA MAZON ALEX MAURICIO	Profesional designado por la máxima autoridad, quien lo presidirá	0803036672	PAZ MORAN VERONICA ALEXANDRA	Titular del área requirente	0803231893	GRACIA CHAMORRO RODOLFO RUBEN	Profesional afín al objeto de la contratación designado por la máxima autoridad	0803102342	BEDON MARIN DORIS MERCEDES	Secretario/a	
Cédula / Núm. Identificación	Apellidos y Nombres	Función en la Comisión Técnica															
1803023827	ARBOLEDA MAZON ALEX MAURICIO	Profesional designado por la máxima autoridad, quien lo presidirá															
0803036672	PAZ MORAN VERONICA ALEXANDRA	Titular del área requirente															
0803231893	GRACIA CHAMORRO RODOLFO RUBEN	Profesional afín al objeto de la contratación designado por la máxima autoridad															
0803102342	BEDON MARIN DORIS MERCEDES	Secretario/a															
Forma de Pago:	Anticipo: 0% Saldo: Pago contra entrega de bienes obras o servicio 100.00%																
Tipo de Adjudicación:	Total																
Plazo de Entrega:	15 días																
Vigencia de Oferta:	60 días																
Funcionario encargado del proceso:	doris.bedon@hdtc.gob.ec																
Estado del Proceso:	Adjudicado - Registro de Contratos																
Descripción:	ADQUISICIÓN DE DISPOSITIVOS MÉDICOS DE LABORATORIOS PARA EL CENTRO DE HEMODIALISIS DEL HOSPITAL GENERAL ESMERALDAS SUR DELFINA TORRES DE CONCHA																
Variación mínima de la Oferta durante la Puja:	1.00% Tipo Variación: Precio total																

Imagen 16: Información extraída del nivel, descripción del proceso

Descripción	Fechas	Productos	Archivos
Fechas de Control del Proceso			SIE-GMCM-036-2020
Fecha de Publicación	2020-10-12 12:00:00	Indicar la fecha real en la cual desea publicar el Proceso.	
Fecha Límite de Preguntas	2020-10-13 12:00:00	Fecha máxima para solicitar aclaraciones respecto al Proceso de Contratación.	
Fecha Límite de Respuestas	2020-10-14 14:00:00	Fecha máxima para solventar cualquier inquietud relacionada al Proceso de Contratación.	
Fecha Límite entrega Ofertas	2020-10-16 14:00:00	Fecha máxima de entrega Ofertas Técnica.	
Fecha Límite solicitar Convalidación	2020-10-20 14:00:00	Fecha máxima para que la Entidad notifique los errores de forma en las ofertas.	
Fecha Límite respuesta Convalidación	2020-10-22 14:00:00	Fecha máxima para respuestas de Convalidación de Errores.	
Fecha Límite de Calificación	2020-10-23 14:00:00	Fecha máxima para calificar a proveedores.	
Fecha Inicio de Puja	2020-10-26 11:00:00	Fecha en que inicia la puja.	
Fecha Final de Puja	2020-10-26 12:00:00	Fecha en que finaliza la puja.	
Fecha de Negociación	2020-10-26 14:00:00	Fecha para la Negociación	
Fecha Estimada de Adjudicación	2020-10-29 16:30:00	Fecha estimada para la Adjudicación de la compra.	

Imagen 17: Información extraída del nivel 1, fechas de control del proceso

Invitaciones a proveedores	
Código:	SIE-CNEZCH-002-2021
Objeto del proceso de contratación:	CONTRATACIÓN DEL SERVICIO DE SEGURIDAD Y VIGILANCIA PARA LA DELEGACIÓN PROVINCIAL ELECTORAL DE ZAMORA CHINCHIPE
Buscar Proveedor:	<input checked="" type="radio"/> Por RUC <input type="radio"/> Por Razón Social
No.	Razón Social - Proveedor
1	ABAD PARDO REYNALDO CESAR
2	ABARCA ABARCA JOSE RAFAEL
3	ABARCA SEGURIDAD ABARCOMSEG CIA.LTDA.
4	ACHAN AQUINO JOSE IGNACIO
5	ACKER S.A.
6	ACUÑA & DELGADO ACUÑAYDELGADO CIA.LTDA.
7	AEROSSEG SERVICIO TECNICO DE SEGURIDAD PRIVADA CIA. LTDA
8	A.E.S.M. ASESORES ESPECIALISTAS EN SEGURIDAD Y MONITOREO CIA. LTDA.
9	AGENCIA DE PROTECTORES PRIVADOS DEL ECUADOR MORAN AGEPROECUA CIA. LTDA.
10	AGENCIA DE SEGURIDAD PRIVADA ASEGPRIV CIA.LTDA.
11	AGENCIA DE SEGURIDAD Y VIGILANCIA PRIVADA ASEVIP CIA.LTDA.

Imagen 18: Información extraída del nivel 2, Invitaciones a proveedores.

Información Esencial del Proceso de Contratación														
Entidad Contratante	INIAP ESTACION EXPERIMENTAL SANTA CATALINA													
Código del Proceso	SIE-INIAPSC-013-2020													
Objeto de Contratación	ADQUISICIÓN E INSTALACIÓN DE UN "SISTEMA ÓPTICO PARA DETECTOR DE FLUORESCENCIA (FLD) DEL EQUIPO DE CROMATOGRAFÍA LÍQUIDA DE ALTA RESOLUCIÓN (HPLC)" DEL LABORATORIO LASTA DEL DEPARTAMENTO NUTRICIÓN Y CALIDAD DE LA ESTACIÓN EXPERIMENTAL SANTA CATALINA													
Tipo de Compra	Bien													
Tipo de Procedimiento	Subasta Inversa Electrónica													
Presupuesto Referencial	\$ 14,867.06													
Anticipo	Porcentaje: 0.00 % Valor: \$ 0.00													
Saldo	Porcentaje: 100.00 % Valor: \$ 14,867.06													
Plazo de entrega	30 días													
Información de Adjudicación														
#	RUC	Nombre del Adjudicatario	Monto de Adjudicación (\$)											
1	1791316290001	SOLUCIONES TECNICAS Y CORPORATIVAS FIELQUIMEC CIA. LTDA.	13,826.37											
Información del Contrato														
#	RUC	Razón Social	Monto Contractual (\$)	Plazo Contractual (días)	Pago Contra Entrega	Renuncia de Anticipo	Anticipo	El Contrato inicia en	Fecha de Suscripción del Contrato	Fecha de notificación de disponibilidad de anticipo	Fecha de otra condición de acuerdo a la naturaleza del contrato	Número de entregas parciales	Fecha Prevista de Terminación de Contrato	Administrador del Contrato Usuario - Nombre
1	1791316290001	SOLUCIONES TECNICAS Y CORPORATIVAS FIELQUIMEC CIA. LTDA.	13,826.37	30	SI	NO	0.00	Fecha de Suscripción de Contrato	2020-11-05	-	-	Entrega Definitiva	2020-12-05	isamaniemgom - IVAN RODRIGO SAMANIEGO MAIGUA

Imagen 19: Información extraída del nivel 2, información esencial del contrato

Módulo de almacenamiento

Se implementó una base de datos relacional mediante el gestor MySQL 8.0.17; la cual es operada por las distintas funciones contenidas en el módulo de almacenamiento. Una vez recibida la información estructurada por parte de la función de extracción correspondiente, el módulo de navegación es el encargado de invocar el método necesario para poder ingresar dentro de la tabla correspondiente de la base de datos la información extraída. En total son once funciones pertenecientes al módulo de almacenamiento, entre las cuales se encuentran métodos de consulta, almacenamiento, conexión y desconexión (ver tabla 3). De suma importancia es la implementación de una relación específicamente para ámbito de almacenamiento de logs generados durante la ejecución del software. En la imagen 20 se indica el esquema de la base de datos propuesta.

Método	Tipo
Connect_to_wpsie_db	Conexión
Dissconnect_db	Desconexión
Get_triad_pks	Consulta
Get_procesos_page_link	Consulta
Get_contrato_links	Consulta
Get_invitacion_links	Consulta
Save_SIE_info	Almacenamiento
Save_contrato_SIE_info	Almacenamiento
Save_invitacion_SIE_info	Almacenamiento
Add_entry_log	Almacenamiento

Tabla 3: Funciones del módulo almacenamiento

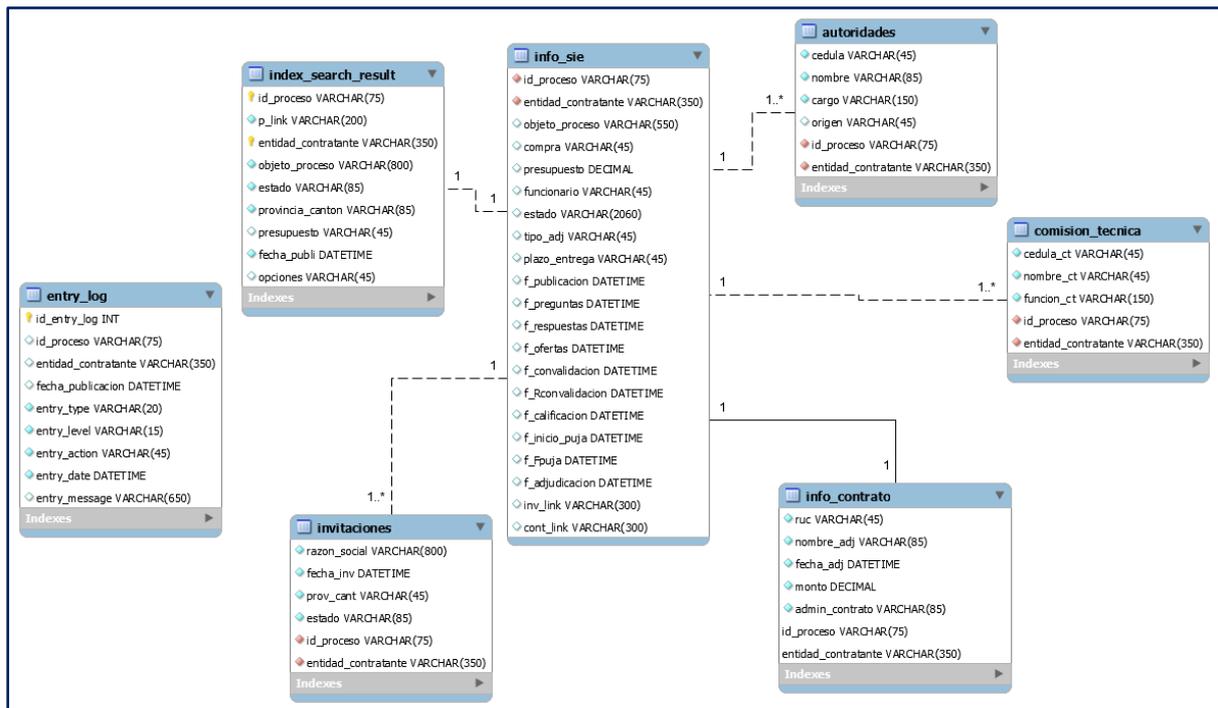


Imagen 20: Esquema de la base de datos

Módulo de utilidades

Dentro de este módulo se encuentran variables de apoyo y configuración para el programa. En total está conformado por 7 variables globales las cuales son: el path hacia el controlador de Google Chrome utilizado por Selenium, la url hacia la página inicial de Búsqueda de Procesos de Contratación de la SERCOP, las variables de fecha de inicio y final que serán utilizadas por el programa para establecer el rango de fechas del cual extraer los datos, una variable para indicar al programa que tipo de configuración de extracción de fechas se va a realizar, ya que el programa brinda la opción de extraer todas las fechas de control del proceso o en caso de requerir un ahorro de memoria se puede establecer la extracción únicamente de las fechas más relevantes, una variable que establece el número límite de invitaciones a recopilar de cada proceso esto debido a que en ciertos casos el número de empresas invitadas a participar puede superar las treinta mil y finalmente se establece un

parámetro para realizar la extracción del nivel 1 y 2 sin necesidad de abrir una ventana del navegador Chrome.

Tecnologías utilizadas

La tabla 4 resume las tecnologías y las versiones utilizadas en el desarrollo de el presente proyecto.

Nombre	Versión	Función
Python	3.6	Lenguaje de programación
Selenium	3.141.0	Automatización de navegador web
MySQL	8.0.17	Gestor de base de datos
Google Chrome	90.0.4430.24	Navegador web
Chrome driver	90.0.4430.24	Controlador para navegador web
PyCharm	2019.3.5	Ambiente de desarrollo

Tabla 4: Tecnologías utilizadas para el desarrollo del proyecto

Mejoras propuestas

En la versión actual del proyecto no se provee un método de solución al CAPTCHA inicial el cual es el único punto necesario de interacción directa de un usuario. En versiones futuras se puede implementar un módulo dedicado a la solución del CAPTCHA lo cual resultaría en la automatización total del programa con lo cual se podría automatizar la ejecución cada cierto periodo de tiempo con el objetivo de mantener la base de datos lo más actualizada posible ante los cambios que se puedan ir dando en el estado de cada proceso y a su vez ir agregando los nuevos procesos. De igual manera se puede ir complementando el módulo de

extracción y navegación con el objetivo de recopilar una mayor cantidad de información y archivos de cada proceso.

El programa actual utiliza el driver para el navegador Google Chrome en la versión 90.0.4430.24 lo cual limita la ejecución del programa a equipos que cuenten con el navegador en dicha versión. Debido a la variabilidad de versiones en las que se puede encontrar el navegador en los diferentes equipos se podría descargar y almacenar los drivers para versiones inferiores y superiores de Chrome y al ejecutar el programa el mismo se encargue de detectar y establecer cual es la versión del driver necesario a utilizar y con ello se flexibilizaría la ejecución del programa.

Debido a la gran cantidad de información que se requiere extraer se podría considerar la opción de implementar multithreading³ a la hora de la ejecución del programa logrando con ello optimizar el tiempo necesario de extracción de la información de los procesos SIE. Todos los datos recabados son públicos y pueden ser de utilidad para trabajos futuros que verifiquen los procesos de contratación pública en busca de irregularidades, por lo cual se podría subir la base de datos a algún repositorio digital para el libre acceso de la comunidad interesada facilitando así sus investigaciones.

Resultados

Para la prueba realizada en este proyecto se estableció como fecha de inicio y fin el 31 de julio de 2020, en el cual constan un total de 71 procesos registrados en diferentes estados de ejecución. Además, se estableció como opción la recopilación de todas las fechas de control

³ Multithreading es una forma de paralelización o de dividir el trabajo en procesos simultáneos. En lugar de dejar una gran carga de trabajo para un único núcleo, los programas con subprocesos dividen el trabajo en múltiples subprocesos de software. Distintos núcleos del CPU procesan estos subprocesos en paralelo para ahorrar tiempo. (Intel, 2021)

del proceso y se delimitó un número de mil invitaciones a proveedores por cada proceso SIE.

Con todo lo dicho previamente se obtuvieron los resultados que se incluyen en la tabla 5.

Tabla	Número de entradas	Tamaño (Mb)
Index_search_result	71	0.06
Info_sie	71	0.09
Autoridades	213	0.07
Comisión técnica	140	0.07
Info_contrato	41	0.03
Invitaciones	60949	17.06
Entry_log	62149	7.51

Tabla 5: Cantidad de datos almacenados en la BD

Con los parámetros indicados previamente el tamaño ocupado en memoria local fue de 24.89 Mb y en tiempo de ejecución alrededor de 4 horas continuas. Como se menciona previamente las invitaciones a proveedores realizadas en cada proceso pueden variar de cientos a miles, de ahí su gran cantidad inclusive la delimitación a un máximo de mil invitaciones por proceso. Es justamente la extracción y almacenamiento de las invitaciones a proveedores la parte del programa que más tiempo demora en finalizar. Cabe señalar que los registros obtenidos corresponden únicamente a un día, en caso de periodos de tiempo más extensos el tiempo de ejecución y la memoria necesaria para almacenamiento puede crecer drásticamente.

CONCLUSIONES

- La implementación y estructura del webcrawler previo al presente trabajo fue tomado como referencia para el desarrollo del proyecto, pese a las diferencias en la estructura de los portales web de procesos de emergencia y procesos de subasta inversa electrónica, se mantuvo los tres módulos principales de navegación, extracción y almacenamiento con el fin de mantener una coherencia entre los dos proyectos y con ello facilitar la conjunción posible a futuro en un proyecto de mayor envergadura que abarque la obtención de datos de diversos tipos de procesos de contratación.
- Los procesos de contratación mediante la modalidad de subasta inversa electrónica (SIE) son utilizados en la adquisición de bienes y servicios normalizados, en donde los proveedores pujan hacia la baja del precio ofertado. Este tipo de contratación ha significado una gran parte del total de compras públicas a lo largo de los años. A pesar de que la SERCOP publica los datos en su portal web, el proceso de obtención y análisis puede resultar tedioso de ahí la necesidad de automatizar la extracción y estructuración de la información de cada proceso en sus diferentes etapas.
- Se logró el desarrollo de un webcrawler implementado en el lenguaje de Python y con la utilización de la tecnología de automatización Selenium, el mismo fue capaz de recopilar y almacenar los diferentes datos acerca de los procesos de contratación SIE en una base de datos local. Debido a ciertos errores en el diseño del portal web de la SERCOP, como la duplicidad de id para diferentes elementos web el webcrawler fue implementado tomando en cuenta la navegación mediante estilos css y direcciones xpath en su mayoría.

- El driver para el navegador Google Chrome utilizado funciona con una versión en específico, por ello en cada nueva actualización del navegador es necesario descargar los drivers correspondientes a la nueva versión con el fin de mantener la operatividad del programa. Se plantea una solución a futuro en la cual se descarguen una serie de drivers para las diferentes versiones de Chrome y el programa sea el encargado de detectar la necesaria para su funcionamiento. De igual manera se recomienda la actualización periódica de las librerías utilizadas en el proyecto para precautelar la operabilidad del programa.
- Con el fin de optimizar el tiempo de ejecución del programa especialmente durante la extracción y almacenamiento de los niveles 1 y 2 de cada proceso, se configuró al driver de selenium con la opción de *headless* lo cual implica que no se mostrará una ventana del navegador Chrome directamente al usuario en su lugar todo el proceso se realizará en background mejorando el tiempo ejecución del programa.
- Toda la información recopilada y almacenada puede ser de utilidad para el desarrollo de futuros proyectos encargados del seguimiento y verificación de las compras públicas realizadas por el gobierno, por lo cual se puede analizar la posibilidad de subir a un repositorio web público toda la base de datos obtenida mediante el presente proyecto.

REFERENCIAS BIBLIOGRÁFICAS

REFERENCIAS

FERNÁNDEZ, Y. (30 de Octubre de 2018). Captcha y reCaptcha: qué son, cuáles son sus diferencias y qué cambia con el reCaptcha v3.

Intel. (2021). Obtenido de ¿Qué es Hyper Threading?

Ionos. (10 de Septiembre de 2020). ¿Qué es el web scraping?

Neoattack. (s/f). ¿Qué es el Crawling?

octoparse. (30 de Diciembre de 2019). ¿Qué Es Web Scraping (Web Crawler) y Cómo Funciona?

SERCOP. (2016). Ecuador.

SERCOP. (21 de Agosto de 2018). LEY ORGANICA DEL SISTEMA NACIONAL DE. Ecuador.

ANEXO A: DOCUMENTACIÓN DEL PROGRAMA

Adjunto al documento se encuentra la documentación de todo el código utilizado en el desarrollo del proyecto. El mismo fue generado de manera automática y se encuentra organizado en un archivo HTML dentro del directorio */docs/build/html* con el nombre *index.html* y se encuentra dentro de la carpeta del proyecto con nombre homónimo: 'WECPSIE'.