

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

**Data Web Scraping de procesos de Contratación Pública de
Emergencia: DAWSCOPE**

Jonathan Patricio Cazco González

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, 10 de Diciembre de 2020

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

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

**Data Web Scraping de procesos de Contratación Pública de
Emergencia: DAWSCOPE**

Jonathan Patricio Cazco González

Nombre del profesor, Título académico

Daniel Riofrío, Ph.D

Quito, 10 de Diciembre de 2020

© 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: Jonathan Patricio Cazco Gonzalez

Código: 136618

Cédula de identidad: 1718655333

Lugar y fecha: Quito, 10 de Diciembre de 2020

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UNPUBLISHED DOCUMENT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

Los Procesos de Contratación Pública de Emergencia son procesos de contratación reducida, controlados por la prontitud de subsanar la emergencia y el presupuesto disponible de la entidad pública contratante. En la última emergencia sanitaria (COVID-19) se registraron irregularidades detectadas gracias a la publicación de estos procesos en el portal web de la SERCOP. Sin embargo, la vasta información y el esquema en la que está organizada dificulta la transparencia de estos datos. El motivo principal de este proyecto es la recopilación automática y sincronizada de todos los datos expuestos a través del uso e implementación de web crawlers (bots). Conseguir una copia reestructurada de estos datos conforma el primer paso en un proyecto de Data Science, el cual busca analizar esta información para monitorear estos procesos y detectar posibles irregularidades. Este trabajo de titulación sirve como una prueba de concepto con la finalidad de determinar tecnologías óptimas para esta tarea, así como la relevancia de los datos extraídos.

Palabras clave: contratación pública, irregularidades, webcrawler, navegación, extracción, almacenamiento, python.

ABSTRACT

The Emergency Public Procurement Processes are reduced procurement processes, ruled by the promptness of assisting the emergency and the available budget of the public contracting entity. In the last health emergency (COVID-19), irregularities were detected thanks to the publication of these processes on the SERCOP web portal. However, the vast information and its organization scheme makes it difficult to determine the transparency of this data. The main reason for this project is the automatic and synchronized collection of all exposed data through the use and implementation of web crawlers (bots). Obtaining a restructured copy of this data is the first step in a Data Science project, which seeks to analyze this information to monitor these processes and detect possible irregularities. This capstone project serves as a proof of concept in order to determine the optimal technologies for this task, as well as the relevance of the extracted data.

Keywords: public procurement, irregularities, webcrawler, browsing, extraction, storage, python.

TABLA DE CONTENIDO

<i>Introducción</i>	7
<i>Antecedentes</i>	7
<i>Objetivos</i>	8
Objetivo General	8
Objetivos Específicos	8
<i>Desarrollo del Tema</i>	9
El Sistema Oficial de Contratación Pública de Emergencia	9
Esquema de Navegación	11
Web Crawler	13
Módulo de navegación (Spider-Dawscope crawler)	15
Módulo de extracción (Extract-Dawscope scrapper)	21
Módulo de almacenamiento (Store-Dawscope database)	23
Módulo de utilidades (Utilities-Dawscope).....	24
Base de datos	24
Resumen de tecnologías utilizadas	25
Resultados	26
<i>Conclusiones</i>	28
<i>Referencias bibliográficas</i>	30
<i>Anexo A: Documentación del código fuente</i>	31
<i>Anexo B: Esquema gráfico de la base de datos</i>	32

ÍNDICE DE TABLAS

Tabla 1. Keywords y enrutamiento a seguir por web crawler.	13
Tabla 2. Relación entre niveles de navegación, métodos y páginas SERCOP.	15
Tabla 3. Clasificación de los métodos de extracción segun su nivel de acción y tipo de información extraída.	21
Tabla 4. Relación entre niveles y tablas en la base de datos.	25
Tabla 5. Tecnologías utilizadas para el desarrollo de DAWSCOPE.	25
Tabla 6. Tamaño de tablas en base de datos de pruebas.	26

ÍNDICE DE FIGURAS

Figura 1. Flowchart del esquema de navegación a seguir.	12
Figura 2. Componentes de DAWSCOPE.	14
Figura 3. Nivel Index de navegación y sus elementos clave.	16
Figura 4. Nivel 0 de navegación y sus elementos clave.	18
Figura 5. Nivel 1 de navegación y sus elementos clave.	19
Figura 6. Nivel 2 de navegación y sus elementos clave.	20
Figura 7. Tipos de tablas en las páginas de navegación.	22
Figura 8. Snippets de código del módulo de almacenamiento.....	24

INTRODUCCIÓN

El presente proyecto propone la recopilación, estructuración y almacenamiento de datos sobre contrataciones públicas de emergencia expuestos en la página web del Servicio Nacional de Contratación Pública (SERCOP). Esta copia sincronizada de los procesos de contratación expuestos permitirá el estudio de esta información de manera que a futuro se pueda detectar cualquier irregularidad en este tipo de procesos. El principal actor encargado de la recolección de datos será un bot de tipo web crawler que sea capaz de navegar la estructura de páginas de cada proceso y extraer información expuesta en tablas y archivos que respalden dicha compra pública. De manera paralela, esta información será estructurada en una base de datos que permita su fácil acceso y manipulación. Finalmente, este proyecto también pretende brindar una abstracción informática de los sistemas de compras públicas en emergencia, exponiendo puntos clave en este tipo de procesos y traduciendo dichas variables en forma de un banco de datos sobre el cual se pueda continuar una investigación.

ANTECEDENTES

Durante los meses de la emergencia sanitaria (COVID-19) se dieron a conocer varios contratos con irregularidades realizados por entidades públicas. Según el noticiero 24 HORAS, desde Marzo a Noviembre del presente año, se han registrado 196 causas penales por delitos de administración pública; la mayoría relacionadas con sobreprecios en la compra de insumos médicos y las redes de corrupción en hospitales públicos (2020). Esto ha generado escepticismo y desagrado en la población ecuatoriana ante la negligencia por parte de las autoridades de control, en este caso la SERCOP, debido a que las denuncias se realizaron de manera externa a las instituciones. Estas denuncias provocaron una reestructuración y auditoría de las entidades

públicas involucradas. Sin embargo, del total de causas solo 13 han pasado a la etapa de instrucción fiscal, mientras que el resto continúan en etapa de indagación previa (24 HORAS, 2020). A pesar de la visión de transparencia del sitio web de la SERCOP al hacer públicos los procesos y documentos de contratación pública, la falta de conocimiento e interés de la ciudadanía y de algún organismo investigativo aparte de los medios de comunicación demanda la búsqueda de una solución automatizada.

OBJETIVOS

Objetivo General

- Recopilar, organizar y almacenar información de dominio público exhibida en la página web de la SERCOP la cual respalda los procesos de compras públicas de emergencia.

Objetivos Específicos

- Entender la estructura del sistema de búsqueda de procesos de contratación pública en emergencia, la información proveída por la página web y los documentos que respaldan dichas actividades.
- Automatizar el proceso de extracción de documentos e información del portal de la SERCOP a través del uso de web crawlers (bots) desarrollados en Python y Selenium.
- Garantizar que el almacenamiento de datos conserve congruencia y actualidad de donde fueron extraídos; además que dicho sistema provea una mejor forma de administrarlos.
- Identificar si los datos recopilados son suficientes para ser sujetos a análisis e investigación posterior.

DESARROLLO DEL TEMA

El Sistema Oficial de Contratación Pública de Emergencia

Las páginas web en las que se presentan los procesos de contratación pública siguen un esquema simple, riguroso pero muy poco intuitivo para alguien que no esté familiarizado con el tema. En un inicio, la página principal (Búsqueda de Publicación de Emergencia) presenta una tabla que cuenta con distintos campos para ingresar filtros de búsqueda para los procesos.

Un CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart: test de Turing público y automático para distinguir a los ordenadores de los humanos) es un tipo de medida de seguridad conocido como autenticación pregunta-respuesta [...] La prueba de un CAPTCHA consta de dos partes simples: una secuencia de letras o números generada aleatoriamente que aparece como una imagen distorsionada y un cuadro de texto. Para superar la prueba y probar que eres un ser humano, simplemente tienes que escribir los caracteres que veas en la imagen del cuadro de texto (Google, 2020).

Una prueba de este tipo debe ser resuelta para poder efectuar la búsqueda deseada y así tener acceso a todos los procesos de contratación de emergencia publicados hasta la fecha de la presente investigación (8714 y en aumento). Los resultados de la búsqueda se muestran en una tabla hasta 20 items, en donde cada proceso es identificado por un código también llamado “Número de Resolución.” Este código contiene el enlace hacia la página de información del proceso (Información de Resolución de Emergencia), donde se expone la información mínima esencial con la que debe contar un proceso de contratación, tal como sus fechas de publicación, las entidades contratantes, las autoridades a su cargo, y documentos que respaldan y motivan

este proceso. Absolutamente todos los procesos de contratación pública tienen representación en la tabla de resultados y su página de información mínima.

Según el avance de la contratación pública, se añaden hasta dos páginas extras para organizar la nueva información sobre el proceso. Bajo el enlace “Ver resultados parciales/finales de las contrataciones realizadas” se encuentra el siguiente nivel de contratación correspondiente a los procesos de adjudicación realizados por la entidad pública. Las contrataciones realizadas se presentan en un formato de tabla y cada una se identifica por el número de resolución seguido de un serial. Finalmente, cada código contiene el enlace para el último nivel de información, en donde se detalla el resumen de dicha contratación. Esto incluye información como: ítems de compra, monto adjudicado, a quien ha sido adjudicado, fechas de contratación, y los documentos que conforman el registro de contratos.

Es importante señalar detalles claves en el ciclo de vida de cada proceso de contratación. Dichos detalles ayudan a determinar el comportamiento del programa y además a entender de mejor manera los procesos de contratación pública de emergencia:

- Los procesos de contratación pública de emergencia son procesos simplificados, es decir, las compras son adjudicadas directamente a proveedores según la necesidad y el presupuesto con la que cuente la entidad contratante. Estos procesos no tienen licitaciones ni concursos por adjudicación de compra, se basan en la disponibilidad y prontitud de los proveedores para atender a la emergencia declarada (SERCOP, 2014, Sección 2).
- La entidad contratante puede ser cualquier institución pública y siempre el encargado de cada proceso es el máximo representante de dicha institución, a menos que esta autoridad realice una delegación de funciones. Esto se ve reflejado en un archivo obligatorio bajo el nombre de “Resolución de delegación de funciones” en

la página de “Informacion Resolución de Emergencia” (SERCOP, 2014, Sección 2).

- Cada proceso de contratación cuenta con tres etapas/estados: declaratoria, contrataciones parciales y emergencia superada. La primera etapa constituye la información básica de publicación de cada proceso, acompañada por el documento de resolución declaratoria de emergencia. La segunda etapa se alcanza una vez que se realiza la primera contratación en este proceso. Estas contrataciones se encuentran en la página bajo el link “Ver resultados parciales de las contrataciones”. Finalmente, cuando el proceso de contratación termina, pasa a ser una “Emergencia Superada.” El link de contrataciones cambiará la palabra “parciales” por “finales” y adicionalmente, la tabla informativa mostrará el presupuesto utilizado en esta contratación, la fecha de finalización de la emergencia y se añadirá un archivo obligatorio bajo el título de “Informe general de resultados.”

Esquema de Navegación

En base al análisis del sistema oficial de contratación pública, la solución de navegación a implementarse en el web crawler es ordinal, en base a la secuencia de páginas web. Cada página web representa un nivel de navegación y de manera iterativa se recolectará la información de todos los procesos en cada nivel antes de pasar al siguiente. Esta decisión fue tomada debido a que cada nivel subsecuente, a excepción del primero, utiliza información extraída por del nivel anterior. Además, así se garantiza la independencia de extracción de información para cada nivel, evitando un sobreconsumo de los recursos expuestos (riesgo de que el bot sea detectado y bloqueado) o un posible cierre de sesión (timeout). De esta forma, El web crawler extraerá y almacenará primero la información mostrada en la tabla índice de búsqueda, además del link

que conduce a la página de cada proceso. A continuación según el esquema de navegación definido, el web crawler deberá realizar las acciones de extracción, almacenamiento y recolección de datos para las páginas: “Información de Resolución de emergencia,” “Información de Contrataciones” y “Resumen de la Contratación,” ya que a diferencia de la página inicial de búsqueda en donde se debe resolver un captcha, acceder a estas tres páginas no requiere de ningún tipo de autenticación, solo el link de navegación correspondiente. La figura 1 presenta el esquema de navegación para dicho agente:

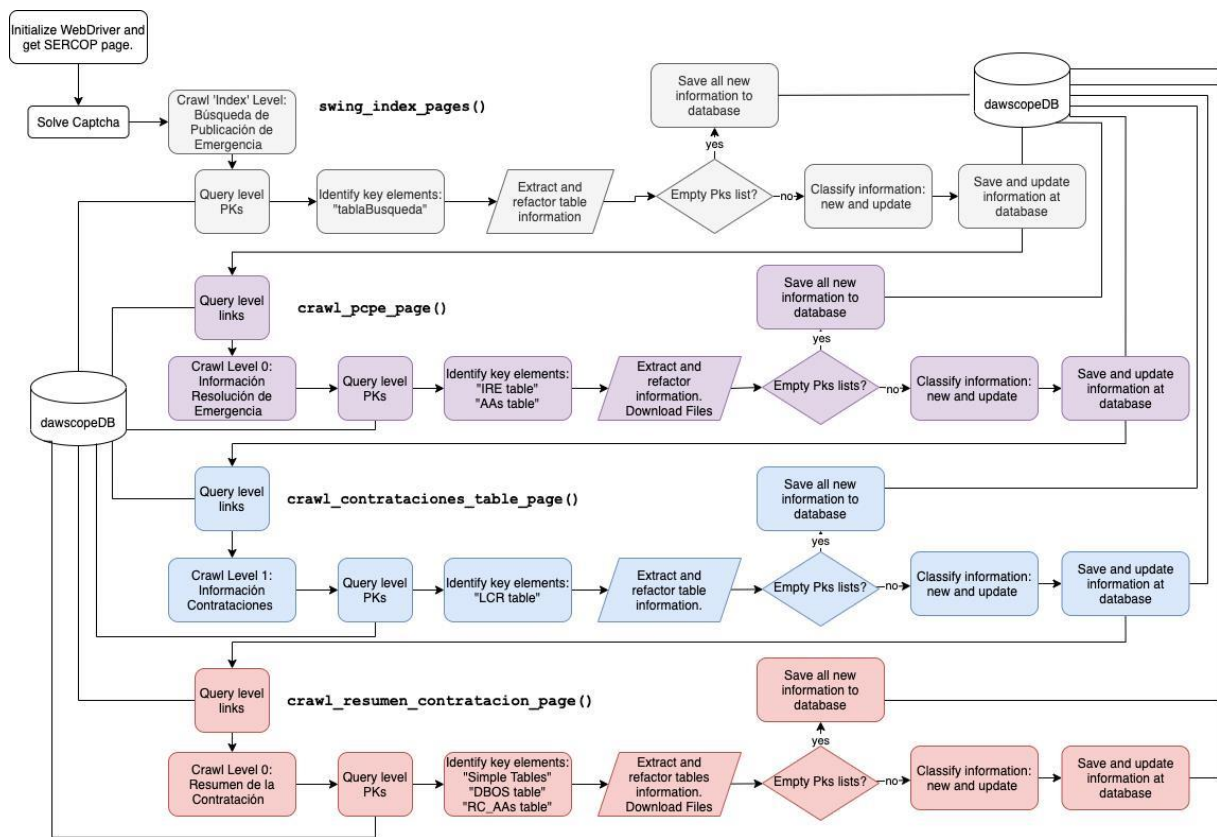


Figura 1. Flowchart del esquema de navegación a seguir.

Para cumplir con las tareas asignadas, el web crawler necesita trazar una ruta a través de las páginas web de cada proceso. Esta ruta está basada en un set de palabras claves que definen sus acciones de exploración. A continuación, la tabla 1 describe este conjunto de palabras clave y el enrutamiento que cada una conlleva:

Origen	Recurso (Keyword)	Tipo (HTML)	Método de localización	Destino
Búsqueda de Publicación de emergencia	Código (Número de resolución) Ver resultados parciales/finales de las contrataciones realizadas	Link_Text	Extracción de celda localizada por ubicación en tabla	Información de Resolución de Emergencia
Información de Resolución de Emergencia	Código (Número de resolución + Serial de Contratación)	Link_Text	Busqueda directa por XPATH	Información Contrataciones
Información Contrataciones		Link_Text	Extracción de celda localizada por ubicación en tabla	Resumen de la Contratación

Tabla 1. Keywords y enrutamiento a seguir por web crawler.

Web Crawler

De acuerdo con la definición del diccionario Merriam-Webster, un web crawler es un programa de computadora que navega páginas web en busca de palabras clave (s.f.).

Construido utilizando el lenguaje de programación Python v3.8 y el módulo para automatizar navegadores web Selenium, el web crawler de este proyecto puede cumplir las tareas de: navegar las paginas web de la SERCOP, extraer información y documentos relacionados con cada proceso de compra pública en emergencia, determinar si la base de datos se encuentra al día con la información extraída, y finalmente guardar dicha información en la base de datos.

Para este proyecto, se planteó la división de trabajo en 3 módulos principales (3 scripts python distintos): navegación, extracción y almacenamiento. El módulo de navegación se encarga de realizar el ‘crawling’ por los distintos niveles y localizar los elementos clave que

contienen información. Además, realiza consultas a la base de datos para determinar la vigencia de los datos almacenados y clasificar la información recolectada. El módulo de extracción se encarga de la recuperación (scrapping) y estructuración de datos de cada elemento clave localizado previamente por el módulo de navegación. Finalmente, el módulo de almacenamiento recibe los datos estructurados para almacenarlos en las tablas correspondientes de la base de datos. La descripción detallada de estos módulos y sus funciones se puede encontrar en el Anexo A de este documento. La figura 2 muestra los componentes de este software con una breve descripción de sus asignaciones principales:

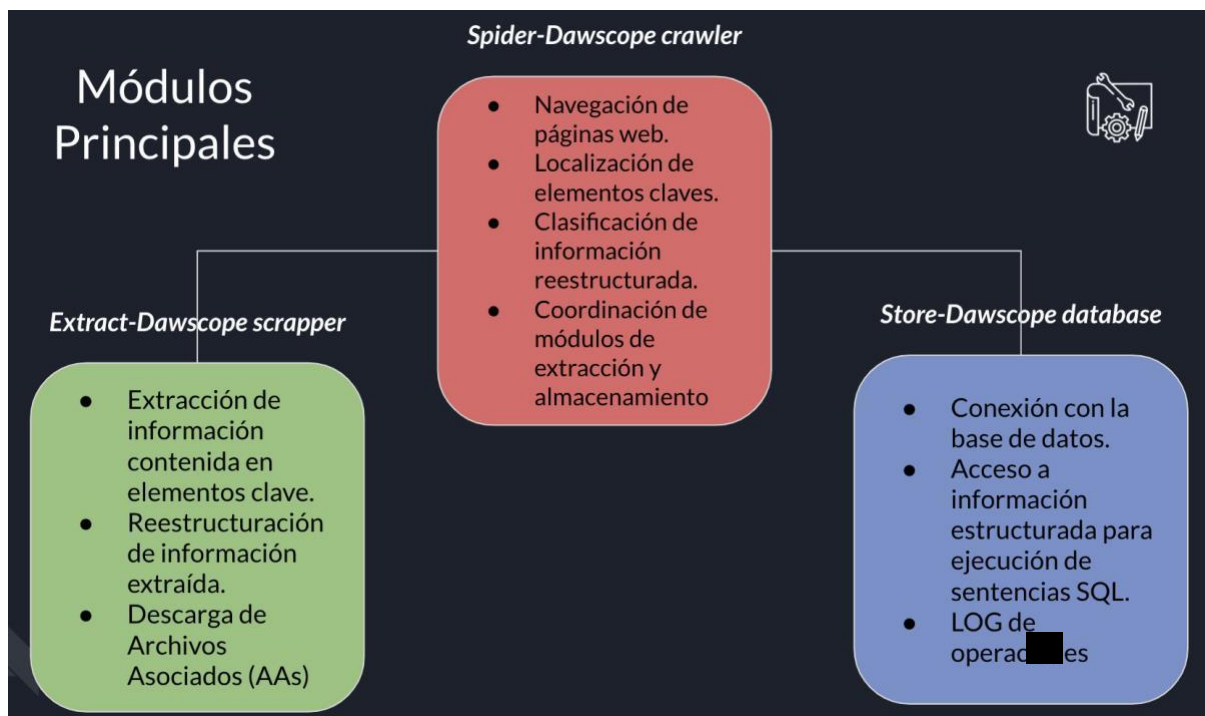


Figura 2. Componentes de DAWSCOPE.

Módulo de navegación (Spider-Dawscope crawler)

El módulo de navegación es el módulo de acción principal del software DAWSCOPE. Es el encargado de abrir y navegar las páginas de cada proceso de contratación pública de emergencia, encontrar los elementos claves en cada una y coordinar las funciones del módulo de extracción y almacenamiento; clasificando correctamente la información extraída para una correcta sincronización con la base de datos. Este módulo cuenta con cuatro métodos principales que corresponden a cada nivel de navegación. Los niveles y métodos definidos para cada página se enlistan en la tabla 2:

Nivel	Método	Página SERCOP
Index	<code>swing_index_pages()</code>	Búsqueda de Publicación de Emergencia
Level 0	<code>crawl_pcpe_page()</code>	Información Resolución de Emergencia
Level 1	<code>crawl_contrataciones_table_page()</code>	Información Contrataciones
Level 2	<code>crawl_resumen_contratacion_page()</code>	Resumen de la Contratación

Tabla 2. Relación entre niveles de navegación, métodos y páginas SERCOP.

El primer nivel de navegación, denominado ‘Index’, contiene 8714 filas (y en aumento) correspondientes a procesos enlistados y publicados en el la tabla índice de búsqueda. Esta tabla índice muestra hasta 20 elementos indexados por página en un total de 436 páginas. Tras la solución del *captcha* y la petición de búsqueda, el crawler comienza por localizar la tabla en donde se muestran los registros gracias al nombre de clase que la define: “tablaBusqueda.” De igual manera, localiza los índices de la página en la que se encuentra, valores que serán importantes para la contabilización de las filas extraídas y para que el programa sepa cuando detenerse. Posterior a eso, se entregan estos elementos al método de extracción `get_main_page_info()`, el cual devuelve una representación estructurada de los datos extraídos para su correcta clasificación. Este método consulta de antemano si es que existe

algún registro previo en la tabla `index_search_result` de la base de datos. Si no consigue información, guarda la información recuperada sin clasificar, invocando el método de almacenamiento `save_index_search_table()`. Caso contrario, clasifica las entradas de nuevos procesos indexados vs procesos ya existentes, e invoca el mismo método de almacenamiento para cada clasificación, alternando el modo entre 'insert' y 'update' respectivamente. Una vez almacenada la información de los 20 procesos en página, el crawler localiza y ejecuta el botón 'Siguiente' gracias a su id en el árbol html: "siguiente-buscar-general," para repetir el mismo proceso hasta completar el número total de procesos indexados hasta la fecha de ejecución del módulo. La figura 3 muestra la página "Búsqueda de Publicación de Emergencia" y la localización de sus elementos clave:

Código	Entidad Contratante	Descripción	Estado	Fecha de Publicación	Opciones
MIDUVI-314-2010	OFICINA TECNICA MIDUVI MANABI	Mediante Res.004 de 23/06/2010,el COE Cantonal Chone resuelve declarar en "Estado de Emergencia" al Proyecto de Reasentamiento Chone, en el afán de evitar una grave conmoción general, por lo que el Sr. Ministro declara en Emergencia tal Proyecto que se ha afectado por incumplimiento del contratista	Cursando Emergencia	2010-07-26 13:13:33	
EMVIAL-050-2010	Empresa Vial del Azuay Envial S.A.	ADQUISICIÓN DE NEUMÁTICOS PARA LA GRUA PROPIEDAD DE LA EMPRESA	Cursando Emergencia	2010-07-26 11:40:12	
050	GOBIERNO AUTONOMO DESCENTRALIZADO ILLUSTRE MUNICIPALIDAD DEL CANTON DAULE	DECLARAR LA EMERGENCIA, SOBRE LA CONTRATACIÓN DE LA CONSTRUCCIÓN DE LAS OBRAS DE COMPLEMENTACIÓN OPERATIVA, SEGURIDAD TÉCNICA Y DURABILIDAD FUNCIONAL DE LA PLANTA DE TRATAMIENTO DE AGUA POTABLE DE LA CIUDAD DE DAULE, EN EL CANTÓN DAULE, DE LA PROVINCIA DEL GUAYAS.	Cursando Emergencia	2010-07-23 15:14:59	
GMCH-001-2010	Municipio El Chaco	DECLARAR EN EMERGENCIA LOS SISTEMAS DE AGUA POTABLE DEL CANTON	Emergencia Superada	2010-07-23 14:41:08	
IML-55	MUNICIPIO DE LOJA	Declarar en emergencia el Sistema de cuenta única del Municipio de Loja, para adquirir los equipos necesarios para su funcionamiento, la instalación de los mismos, en forma directa e inmediata.	Emergencia Superada	2010-07-22 12:41:55	Migrado de Contranet
EME-002-2010	Empresa de Agua Pótale y Alcantarillado EMAPA Regional La Estancilla	RESUELVE: 1° Declarar LA EMERGENCIA anta la falta de unidades sumergibles para la captación de agua cruda que afecta el normal abastecimiento a todas la población servida por la EMAPA REGIONAL LA ESTANCILLA al amparo de lo que dispone el numeral 31 del Art. 6 y el Art. 57 de la Ley Orgánica del	Emergencia Superada	2010-07-22 12:33:07	

Figura 3. Nivel Index de navegación y sus elementos clave.

El segundo nivel de navegación (Level 0) contiene la página obligatoria con la que cuenta cada proceso publicado: Información Resolución de Emergencia. Para ingresar a dicha página, el crawler realiza una consulta de todos los links de procesos almacenados previamente y de manera iterativa abre cada link en el navegador para extraer la información. La peculiaridad de

esta página es que todo su contenido se encuentra dentro de una tabla, haciendo mas fácil la localización de la misma mediante XPATH. El crawler entrega esta tabla a los métodos de extracción `extract_IRe_lv0()`, `extract_AAs_info_lv0()`, `download_AAs()` y recibe de vuelta la información de las tablas “Publicacion de Emergencia” y “Fechas de Control” así como los Archivos Asociados que han sido publicados. Finalmente, se verifica la existencia de la información de las tablas mencionadas en la tabla *info_res_emergencia* de la base de datos y se decide por si hacer una actualización o inserción de los datos. Por su parte, la existencia de los Archivos Asociados también se verifica en la tabla *aas_res_emergencia* gracias a un *hash* de identificación obtenido del *binary-string* del documento descargado.

Una función criptográfica hash- usualmente conocida como “hash”- es un algoritmo matemático que transforma cualquier bloque arbitrario de datos en una nueva serie de caracteres con una longitud fija. Independientemente de la longitud de los datos de entrada, el valor hash de salida tendrá siempre la misma longitud (Donohue, 2014).

En este caso, solo se almacenan los nuevos archivos en la base de datos. Finalmente, este nivel se vale de un solo método de almacenamiento para la información extraída:

save_IRE_and_AAs()). La figura 4 muestra la página “Información Resolución de Emergencia” y la localización de sus elementos clave:

»Información Resolución de Emergencia

[Ver resultados parciales de las contrataciones](#)

Publicación de Emergencia

Entidad Contratante	CORPORACION ELECTRICA DEL ECUADOR CELEC EP
Número de Resolución	RE-CELECEPGG-TPI-102
Descripción de Resolución / Resuelve	Declarar como Régimen de Emergencia la contratación de "Desmontaje, traslado, montaje y puesta en marcha de un Grupo Termoeléctrico de 2,5 MW" a cargo de la Unidad de Negocio CELEC EP-Termopichincha.
Causa de Emergencia	Fuerza Mayor o caso Fortuito
Motivo/Justificación	La situación energética en la Provincia de Esmeraldas es extremadamente complicada y crítica por la coincidencia de la salida de operación de la Central Esmeraldas y el mantenimiento mayor de la Unidad No. 1 de la Propicia.
Tipo Autoridad	Delegado Responsable
Máxima Autoridad o Delegado	JUAN CARLOS LOPEZ
Cargo	GERENTE UNIDAD DE NEGOCIO CELEC EP-TERMOPICHINCHA
Estado de Emergencia	Cursando Emergencia

Fechas de Control

Fecha de publicación de Resolución	2010-07-16 10:27:46
Fecha de Inicio de Emergencia	2010-07-17 06:00:00
Fecha estimada de Finalización de la Emergencia	2010-08-15 20:00:00

Archivos Asociados

Archivo Obligatorio: Resolución Motivada para la declaración de emergencia

Descripción del Archivo	Descargar Archivo
RESOLUCION DE EMERGENICA	Descargar

Documentación Adicional

Archivo Obligatorio: Resolución de delegación de funciones

Descripción del Archivo	Descargar Archivo
NOMBRAMIENTO	Descargar

Figura 4. Nivel 0 de navegación y sus elementos clave.

El tercer nivel de navegación (Level 1) contiene la lista de los procesos de contratación/adjudicación realizados en un proceso de contratación pública de emergencia. Primero se recuperan los links previamente almacenados en la base de datos e iterativamente se despliega cada página en el navegador. El crawler localiza la pequeña tabla informativa la cual contiene el nombre de la entidad contratante, así como el código identificador del proceso. Posteriormente, localiza a través de su XPATH la tabla en donde están indexados los Procesos de Adjudicación (PDAs). A continuación, se entrega esta tabla al método de extracción `extract_LCR_1v11()` el cual devuelve la lista de PDAs encontrados. Esta lista también es clasificada entre registros ‘nuevos’ y ‘preexistentes’ para su correcto almacenamiento en la base de datos, invocando al método de almacenamiento `save_LCR()` con el modo

correspondiente (insert/update). La figura 5 muestra la página “Información Contrataciones” y la localización de sus elementos clave:

Viernes 11 de Diciembre de 2020 | 18:33 [Ingresar al Sistema]

»Información Contrataciones.

La Entidad Contratante aún no ha finalizado la publicación de todas las contrataciones realizadas en esta Emergencia.

Entidad Contratante: JUNTAPARROQUIALDE OYACACHI
Número de Resolución: GADPR-O-001-2020

Listado de Contrataciones Realizadas

Código	Descripción del Proceso de Adjudicación	Provincia/Cantón de Creación	Fecha de Creación	Estado	Acción
GADPR-O-001-2020-00001	ADQUISICIÓN DE MATERIALES DE ASEO QUE PERMITAN EVITAR EL CONTAGIO DEL COVID-19, PARA FUNCIONARIOS DEL GAD PARROQUIAL Y POBLACIÓN DE LA PARROQUIA DE OYACACHI, CON LA FINALIDAD DE SALVAGUARDAR SU INTEGRIDAD FÍSICA, EN RELACIÓN A LA PANDEMIA MUNDIAL POR COVID 19? ENTRE EL GOBIERNO AUTÓNOMO DESCENTRALIZADO DEL GAD PARROQUIAL DE OYACACHI Y LA EMPRESA ECUADORIAN INDUSTRIAL CORPORATION	NAPO/EL CHACO	2020-09-11	Finalizada	
GADPR-O-001-2020-00002	ADQUISICIÓN DE MATERIALES DE ASEO QUE PERMITAN EVITAR EL CONTAGIO DEL COVID-19, PARA LOS FUNCIONARIOS Y SU POBLACIÓN DEL GAD PARROQUIAL DE OYACACHI, CON LA FINALIDAD DE SALVAGUARDAR SU INTEGRIDAD FÍSICA, EN RELACIÓN A LA PANDEMIA MUNDIAL POR COVID 19? ENTRE EL GOBIERNO AUTÓNOMO DESCENTRALIZADO DEL GAD PARROQUIAL DE OYACACHI Y LA EMPRESA ANNA KVIITNA.	NAPO/EL CHACO	2020-09-11	Finalizada	
GADPR-O-001-2020-00005	ADQUISICIÓN DE MATERIALES DE ASEO QUE PERMITAN EVITAR EL CONTAGIO DEL COVID-19, PARA FUNCIONARIOS DEL GAD PARROQUIAL Y POBLACIÓN DE LA PARROQUIA DE OYACACHI, CON LA FINALIDAD DE SALVAGUARDAR SU INTEGRIDAD FÍSICA, EN RELACIÓN A LA PANDEMIA MUNDIAL POR COVID 19? ENTRE EL GOBIERNO AUTÓNOMO DESCENTRALIZADO DEL GAD PARROQUIAL DE OYACACHI Y LA EMPRESA UNI LIMPIO	NAPO/EL CHACO	2020-09-11	Finalizada	

Figura 5. Nivel 1 de navegación y sus elementos clave.

Finalmente, el último nivel (Level 2) representa los detalles de cada contratación realizada. Este nivel es el que más información contiene por página en relación a los niveles posteriores. Tras recuperar los links a cada PDA de la base de datos, el crawler identifica cinco elementos clave a extraer en cada página: “Información de Proceso de Contratación,” “Adjudicación,” “Contratación,” “Detalle: Bienes / Obras /Servicios” y “Registro de Contratos: Archivos Asociados.” Los tres primeros elementos, denominados ‘Simple Tables’ son extraídos utilizando la función `extract_ST_info()` debido a que su estructura es similar a tablas de tipo “Título:Contenido” recurrentes en la plataforma de la SERCOP. La tabla “Detalle: Bienes / Obras /Servicios” (DBOS) es procesada por el método `extract_DBOS_info()`, mientras que los archivos asociados (RC_AAs) son extraídos siguiendo el modelo definido en el nivel 0; en este caso, utilizando las funciones `extract_RC_AA_info_lvl2()` y `download_AAs()`. La clasificación de información para las ‘Simple Tables’ y ‘RC_AAs’ también se realiza de forma similar a la del nivel 0. Por otra parte, la clasificación de la información de la tabla DBOS se realiza a través de la verificación en base de datos del *hash* de cada fila. Este *hash* es calculado utilizando la información extraída de cada fila y el código del respectivo PDA. Finalmente, el

almacenamiento de la información extraída es manejada por los métodos: `save_STs_info()`, `save_DBOS_info()` y `save_RC_AAs()`. A diferencia de los demás niveles, la invocación de métodos de almacenamiento se realiza inmediatamente después de la clasificación de cada elemento (STs, DBOS, RC_AAs) y no después de haber recuperado la información de toda la página. Esto se debe a que la página contiene gran cantidad de información que puede demorar en cargar, lo que ocasionaría que alguno de los elementos clave no sea procesado correctamente. La figura 6 muestra la página “Resumen de la Contratación” y la localización de sus elementos clave:

Viernes 11 de Diciembre de 2020 | 18:36 [Ingresar al Sistema]

» Resumen de la Contratación

Información de Proceso de Contratación

Entidad Contratante	JUNTAPARROQUIALDEOYACACHI
Código	GADPR-O-001-2020-00001
Tipo de Compra	Bien
Provincia / Ciudad	NAPO/EL CHACO
Forma de Pago	Anticipo: 0%; Solicitud Contra Entrega: 100.00%; Saldo a: Pago contra entrega de bienes obras o servicio
Descripción del proceso de adjudicación	ADQUISICIÓN DE MATERIALES DE ASEO QUE PERMITAN EVITAR EL CONTAGIO DEL COVID-19, PARA FUNCIONARIOS DEL GAD PARROQUIAL Y POBLACIÓN DE LA PARROQUIA DE OYACACHI, CON LA FINALIDAD DE SALVAGUARDAR SU INTEGRIDAD FÍSICA, EN RELACIÓN A LA PANDEMIA MUNDIAL POR COVID 19 ENTRE EL GOBIERNO AUTÓNOMO DESCENTRALIZADO DEL GAD PARROQUIAL DE OYACACHI Y LA EMPRESA ECUADORIAN INDUSTRIAL CORPORATION

Detalle: Bienes / Obras / Servicios

Categoría	Bien/Obra/Servicio	Cant.	Unidad Medida	Precio Ref. Unitario	SubTotal
271900931	MASCARILLAS DE PROTECCION (RESOLUCION MOTIVA POR BROTE DE CORONA VIRUS COVID-19)	1	Unidad	USD 2,323.89	USD 2,323.89
Total:					USD 2,323.89

Adjudicación

RUC proveedor adjudicado	1710939925001
Razón social	CHILIGUANO ALMEIDA GLADYS JANETH
Motivo de la adjudicación	POR CONVENIR A LOS REQUERIMIENTOS DE LA INSTITUCION
Monto adjudicado	USD 2,323.89

Contratación

Fecha de inicio del Objeto del proceso de Contratación (Contrato)	2020-04-02
Fecha de término del Objeto del proceso de Contratación(Contrato)	2020-04-06
Monto del Objeto de Contratación(Contrato)(USD)	USD 2,323.89

Archivos Asociados

Registro de contratos

Descripción del Archivo	Descargar Archivo
CONTRATO ECUADORIAN-GLADYS CHILIGUANO	 Descargar

[Regresar](#) [Imprimir](#)

Figura 6. Nivel 2 de navegación y sus elementos clave.

Módulo de extracción (Extract-Dawscope scrapper)

Este módulo se encarga de la extracción y estructuración de datos recopilados a partir de los elementos web entregados por el módulo de navegación. Cuenta con ocho métodos diseñados en base a las necesidades de extracción según el tipo de objeto recibido. La tarea de estos métodos también consiste en organizar la información extraída en estructuras de datos (en su mayoría listas de diccionarios estilo JSON) para poder manipular fácilmente dicha información al momento de clasificarla y almacenarla en la base de datos. La clasificación de estos métodos según el nivel desde donde son invocados y el tipo de información extraída se encuentra en la tabla 3:


Nivel	Método	Objeto Web analizado	Información extraída
Index	get_main_page_table_info()	Tabla: Búsqueda	Texto, Links
Level 0	extract_IRE_lv0()	Tabla: IRE	Texto, Links
	extract_AAs_info_lv0()	Tablas: IRE_AAs	Texto, Links
Level 0, Level 2	download_AAs()	Lista de archivos presentes.	Archivos
Level 1	extract_LCR_lv11()	Tabla: LCR	Texto, Links
		Tablas: IPC,	Texto
Level 2	extract_ST_info()	Adjudicación, Contratos	
	extract_DBOS_info()	Tabla: DBOS	Texto
	extract_RC_AA_info_lv12()	Tabla: RC_AAs	Texto, Links

Tabla 3. Clasificación de los métodos de extracción según su nivel de acción y tipo de información extraída.

Los métodos para extracción de texto en este módulo utilizan mecanismos iterativos sobre tres tipos de tablas comunes para exponer información en las distintas páginas: tablas $2 \times n$ de tipo "Titulo:Contenido", tablas $2 \times n$ de tipo "Titulo:Archivo" presentes en los niveles 0 y 2, y tablas $n \times m$ tipo matriz con n columnas como categorías y m filas como instancias de un tipo

de dato relacionado con el nivel en el que se encuentra. La figura 7 ayuda a ejemplificar de mejor manera la descripción de estas tablas:

1.	Entidad Contratante	JUNTAPARROQUIALDEOYACAHÍ
	Código	GADPR-O-001-2020-00001
	Tipo de Compra	Bien
	Provincia / Ciudad	NAPO/EL CHACO
	Forma de Pago	Anticipo: 0%; Solicitud Contra Entrega: 100.00%; Saldo a: Pago contra entrega de bienes ob
Descripción del proceso de adjudicación		ADQUISICIÓN DE MATERIALES DE ASEO Q DE SALVAGUARDAR SU INTEGRIDAD FÍSIC ECUADORIAN INDUSTRIAL CORPORATION

2.	Descripción del Archivo	Descargar Archivo
	RESOLUCION MOTIVA POR BROTE DE CORONA VIRUS COVID-19	 Descargar

Código	Entidad Contratante	Descripción	Estado	Fecha de Publicación	Opciones
004-ALC-GADM/L-2020	Gobierno Autónomo Descentralizado Municipal de Limón Indanza	RATIFICAR la declaratoria del estado de emergencia del Relleno Sanitario del Cantón Limón Indanza, en virtud de los informes técnicos presentados sobre el calamitoso estado de la celda.	Cursando Emergencia	2020-12-10 17:25:37	
1302-AC-GADPMS-2020	GAD DE LA PROVINCIA DE MORONA SANTIAGO	DECLARAR EN EMERGENCIA LA CONSTRUCCIÓN DEL PUENTE SOBRE EL RÍO PANANZA SECTOR SANTA ROSA DE ORO, UBICADO EN LA PARROQUIA SANTIAGO DE PANANZA, DEL CANTÓN SAN JUAN BOSCO	Cursando Emergencia	2020-12-04 23:20:42	Migrado de Contratanet
1301-GADPMS-2020	GAD DE LA PROVINCIA DE MORONA SANTIAGO	DECLARACION DE EMERGENCIA PARA LA CONSTRUCCION DEL PUENTE SOBRE EL RO ZAMORA PASO A LAS COMUNIDADES DE EL ARENAL SAN JOSE DE PUNTO AVANTAS CAMPANA E.T.S.A CONDOR Y JAIME NARVAEZ DE LA PARROQUIA BOMBOZA CANTON GUALAQUIZA	Cursando Emergencia	2020-12-04 23:11:40	Migrado de Contratanet
GADMFA-05-2020	GOBIERNO MUNICIPAL DE FLAVIO ALFARO	DECLARATORIA DEL ESTADO DE EMERGENCIA POR EL GRAVEPROBLEMA DE LA SEQUÍA EN EL CANTÓN FLAVIO ALFARO	Cursando Emergencia	2020-12-03 20:12:34	

Figura 7. Tipos de tablas en las páginas de navegación.

Por otra parte, la extracción y descarga de documentos de esta plataforma se realiza utilizando *multithreading*. Esta es la razón de que existan dos pasos para la descarga de archivos: la recolección de información y el link de cada archivo, y la descarga del mismo. Una vez obtenida la lista de información sobre los archivos publicados, el método `download_AAs()` ordena simultáneamente la descarga de todos los archivos contenidos en dicha lista a través de peticiones HTTP (GET REQUEST). Debido a que la petición devuelve un archivo binario, se utiliza el módulo *magic* el cual analiza los ya conocidos ‘magic numbers’ (2048 bits iniciales) para determinar el tipo/extensión del archivo descargado.

Finalmente, este módulo también se vale de técnicas de hashing para asignar uniqueness en información sin identificador. Utilizando el módulo de *hashing, blake2*, se genera un

identificador de 40 caracteres basado en componentes ‘únicos’ de cada tipo de dato. En este caso los Archivos Asociados llevan un identificador obtenido de la representación binaria del propio archivo (*bin-string*), mientras que los ítems de compra en la tabla DBOS llevan un identificador basado en la combinación de la información de sus columnas más el id del proceso de adjudicación (PDA) al que pertenece.

Módulo de almacenamiento (Store-Dawscope database)

El módulo de almacenamiento es el encargado de manejar todas las transacciones con la base de datos MySQL gracias al módulo/driver *mysql.connector*; el cual encapsula tanto la conexión con la base de datos como la ejecución de sentencias SQL. Store-Dawscope database cuenta con 19 métodos diseñados para conectar los datos estructurados con la abstracción de tablas de la base de datos. De acuerdo a su funcionalidad, estos métodos pueden ser clasificados en 4 categorías: conexión, consulta, almacenamiento y log.

La categoría de conexión cuenta con dos métodos los cuales se encargan de establecer/terminar la conexión con la base de datos. Los demás métodos dependen de que esta conexión esté abierta para realizar sus transacciones. La categoría de consulta contiene diez métodos que realizan consultas sobre la base de datos. Todos estos métodos recuperan información sobre Primary Keys (PKs) de distintas tablas para que el módulo de navegación clasifique correctamente la información nueva y por actualizar. La categoría de almacenamiento cuenta con seis métodos que se encargan de acceder la información de los diccionarios entregados, emparejarla con el *query* respectivo (inserción/actualización) y ejecutar la sentencia en la base de datos. Finalmente, el método de log es uno de los más importantes en todo el software Dawscope. Con un total de 26 referencias, este método se encarga de registrar el funcionamiento del software, documentando cualquier éxito/fallo de

navegación, extracción o almacenamiento en la ejecución del programa. La figura 8 muestra una *snippet* de código representativo por cada categoría descrita:

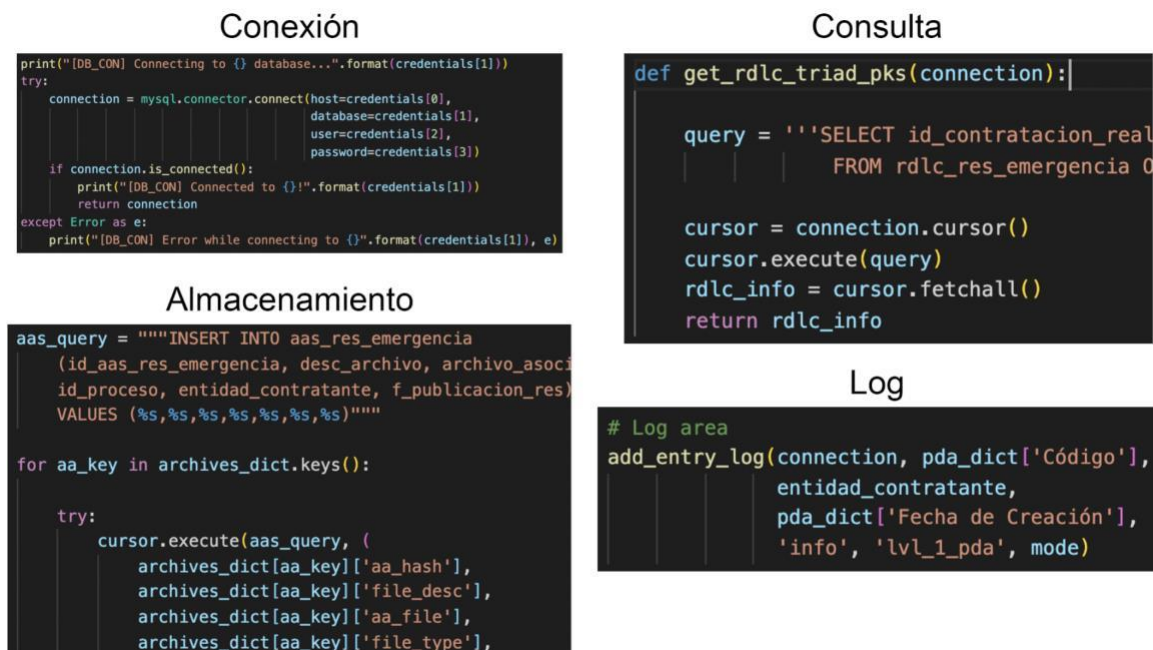


Figura 8. Snippets de código del módulo de almacenamiento.

Módulo de utilidades (Utilities-Dawscope)

Este es el módulo más pequeño de todo el programa. Cuenta con apenas dos métodos y dos variables globales. El objetivo de este módulo es encapsular y reutilizar tareas pequeñas que son de gran utilidad para los módulos principales. Las variables globales son: el *path* hacia el controlador de Google Chrome utilizado por Selenium y el url inicial a la página de “Búsqueda de Publicación de Emergencia” de la SERCOP.

Base de datos

La base de datos fue desarrollada utilizando el gestor de bases de datos relacionales MySQL, version 8.0.18. Las tablas fueron modeladas en base a cómo la información está esquematizada en los distintos niveles, procurando guardar la cohesión e integridad de datos

extraídos y a su vez proveer una mejor manera de administrarlos. Esta base de datos cuenta con 7 tablas relacionadas que almacenan la información de cada P.C.P.E y una tabla adicional para registrar el funcionamiento del software (LOG). El esquema gráfico completo de esta base de datos se encuentra en el Anexo B. La tabla 4 muestra la relación de cada tabla/entidad en la base de datos con su respectivo nivel y el tipo de información que almacena:

Nivel	Tabla	Información Almacenada
Index	<i>index_search_result</i>	Texto, Links
	<i>info_res_emergencia</i>	Texto, Links
Level 0	<i>aas_res_emergencia</i>	Archivos, Texto
Level 1	<i>lcr_res_emergencia</i>	Texto, Links
	<i>rdlc_res_emergencia</i>	Texto
Level 2	<i>dbos_res_emergencia</i>	Texto
	<i>rc_aas_res_emergencia</i>	Archivos, Texto
Log	<i>entry_log</i>	Logs (Texto)

Tabla 4. Relación entre niveles y tablas en la base de datos.

Resumen de tecnologías utilizadas

La tabla 5 muestra las tecnologías utilizadas para el desarrollo de este proyecto:

Nombre	Versión	Función
Python	3.8.6	Lenguaje de desarrollo
Selenium	3.14.0	Automatizador de navegadores web
MySQL	8.0.18	Base de datos relacional
Google Chrome	87.0.4	Navegador web
ChromeDriver	87.0.4	Controlador de navegador web
Visual Studio Code	1.52.1	Ambiente de desarrollo (IDE)

Tabla 5. Tecnologías utilizadas para el desarrollo de DAWSCOPE.

Resultados

- La base de datos de pruebas de la versión 1.0 de este software cuenta con la información extraída de todos los niveles de 100 a 150 procesos de contratación, además de 13108 entradas en la tabla de LOG, las cuales registran el trabajo y los errores del software a lo largo de su desarrollo. Esta base de datos tiene aproximadamente 1.4Gb de información. La tabla 6 muestra el tamaño en *Megabytes* de cada tabla en la base de datos de prueba:

Tabla	Tamaño (MB)
<i>aas_res_emergencia</i>	666.61
<i>rc_aas_res_emergencia</i>	648.59
<i>info_res_emergencia</i>	6.55
<i>index_search_result</i>	4.55
<i>entry_log</i>	2.52
<i>lcr_res_emergencia</i>	0.23
<i>rdlc_res_emergencia</i>	0.13
<i>dbos_res_emergencia</i>	0.09

Tabla 6. Tamaño de tablas en base de datos de pruebas.

- Los ensayos de manipulación, transportación y duplicación de estos datos a la base de datos de pruebas del proyecto permitieron evidenciar algunas falencias del sistema de la SERCOP como la duplicación de códigos identificadores únicos para cada proceso o la existencia de procesos que pese al vencimiento de su fecha estimada de finalización, siguen constando como abiertos ('cursando emergencia') en el portal web.
- Un ensayo de recolección completo de información de todos los procesos sobre los dos primeros niveles (Index, Level 0) dió como resultado un tiempo de ejecución de 11 horas con 30 minutos. Además, el tamaño de la base de datos generada llegó a ser de aproximadamente 80Gb. Tras un análisis de estos resultados y las pruebas de 100 a 200 iteraciones/procesos en los niveles subsecuentes, se extrapoló que el tiempo total aproximado de ejecución para todos los niveles será de 25 a 35 horas, con una base de datos

de 350-375 Gb aprox. Los tiempos de ejecución podrían reducirse o aumentar en base al ancho de banda de la red en la que corra este programa (50Mbps para este trabajo de titulación).

Errores Conocidos y Estudios Adicionales

- Debido al uso de métodos iterativos para extraer la información, cualquier página que falle o demore en la carga de datos puede hacer fallar al crawler por la diferencia de tiempos entre la carga de la página web y la espera del crawler antes de buscar los elementos clave. La forma en la que el software está implementado informa sobre este suceso en la tabla de LOG pero no provee un método de recuperación inmediato. Esto se debe a que la primera versión de este software está enfocada en obtener la mayor cantidad de información posible (greedy algorithm) exceptuando posibles fallos de carga de datos. La única solución a este problema es el volver a correr la aplicación, ya que los protocolos insert/update parcharían cualquier información faltante.
- La versión 1.0 de este software no provee un módulo para la solución automática del *captcha* inicial de búsqueda, siendo el único punto de intervención necesario del usuario con el programa. Versiones posteriores del software requieren la implementación de un módulo de este estilo, el cual utilice machine learning y procesamiento de imágenes. El código fuente de esta versión incluye un módulo que intentó resolver este inconveniente, pero el tipo de *captcha* presentado en esta página resultó ser diferente al método de resolución escogido.
- La implementación de bloques try/except en los métodos de navegación se realiza de forma general, es decir todo el código dentro de un mismo bloque. Esto podría mejorarse aumentando los bloques try/except por cada función de localización o extracción. A la par,

el aumento de estos bloques supondría un mayor dominio en el manejo de excepciones y clases de errores relacionados con el módulo de automatización Selenium WebDriver.

- Uno de los mayores riesgos al que se enfrenta este proyecto sería la modificación de la interfaz gráfica de la página proveedora de información. En un principio, esto modificaría la localización ya configurada de la información que se desea extraer, ocasionando que el agente extraiga información errada o no existente. De igual manera, cualquier cambio en la estructura interna de este servicio provisto por la SERCOP significaría a una re-parametrización del agente para que este se adapte a las nuevas políticas implementadas por dicha institución.

CONCLUSIONES

- Los procesos de contratación pública de emergencia (PCPE) son procesos reducidos de adjudicación, los cuales se rigen por la prontitud de los proveedores para suplir la emergencia y la disponibilidad de presupuesto asignado a la entidad contratante. Cada PCPE tiene tres etapas: declaratoria, contrataciones parciales y emergencia superada; y su información está distribuida en 4 niveles/páginas web: “Búsqueda de Publicación de Emergencia,” “Información Resolución de Emergencia,” “Información Contrataciones” y “Resumen de la Contratación.” La SERCOP utiliza este servicio web como un repositorio digital en donde se expone la información y el desarrollo de cada proceso, así como para asegurar la transparencia este tipo especial de contrataciones.
- La extracción de información de este portal web se logró gracias a la implementación de un WebCrawler que utiliza la tecnología de automatización de navegadores web Selenium. Las decisiones de navegación de este bot fueron diseñadas en base a recursos web como clases de estilo y xpath, los cuales permiten la localización y extracción de la información deseada. Además se requirió de una previa familiarización de los desarrolladores con el

portal web de la SERCOP, su organización y las relaciones entre los distintos niveles de navegación. Siguiendo el esquema de presentación de las páginas web, el web crawler procesa todos los PCPE iterativamente garantizando la extracción de la mayor cantidad de datos posible nivel a nivel.

- La solución de sincronización para todos los módulos consiste en la inserción de nuevos registros y la sobrescritura de registros existentes. Este método de actualización se debe a que la información publicada en los distintos niveles de cada proceso no puede ser modificada. Es decir, en el ciclo de vida de cada proceso, la entidad contratante solo puede seguir añadiendo información en cada nivel de sus respectivos procesos, sin modificar la que ya fue previamente publicada.
- La reestructuración y almacenamiento de la información extraída da como resultado un modelo que permite el análisis en conjunto de todos los niveles de cada proceso de contratación. El esquema de base de datos entregado relaciona la información extraída en todos los niveles y constituye el primer paso en un proyecto de Data Science: la recolección y modelamiento de la información.
- La implementación de bots para recolección de información pública representa una gran posibilidad para entidades destinadas al seguimiento y verificación de procesos efectuados por entidades públicas de cualquier índole, ya que simplifica la minuciosidad de búsqueda específica y da paso a una perspectiva global en cualquier sistema de repositorio digital del estado. Con una suite de software conformado por un recolector de información como el de este trabajo de titulación y otro programa destinado al análisis de los datos recolectados, casos como la denuncia de sobrepagos en el Hospital Teodoro Maldonado Carbo podrían ser detectados y tener un respaldo documentado antes de cualquier modificación conveniente en el repositorio digital de origen.

REFERENCIAS BIBLIOGRÁFICAS

- 24 HORAS. *Se han registrado 196 causas penales por delitos de administración pública* [Archivo de video]. (2020, Diciembre 15). Extraído de <https://www.teleamazonas.com/se-han-registrado-196-causas-penales-por-delitos-de-administracion-publica/>
- Donohue., B. (10 Abr 2014). *¿Qué Es Un Hash Y Cómo Funciona?*. Kaspersky daily. Extraído de <https://latam.kaspersky.com/blog/que-es-un-hash-y-como-funciona/2806/#:~:text=10%20Abr%202014-,Una%20funci%C3%B3n%20criptogr%C3%A1fica%20hash%2D%20usualmente%20conocida%20como%20%E2%80%9Chash%E2%80%9D%2D,tendr%C3%A1%20siempre%20la%20misma%20longitud.>
- Google. (2020). *¿Qué es un CAPTCHA?* Ayuda de Administrador de Google Workspace. Extraído de <https://support.google.com/a/answer/1217728?hl=es>
- Merriam-Webster. (s.f.). Web crawler. In *Merriam-Webster.com dictionary*. Extraído el 24 de Septiembre, 2020, de <https://www.merriam-webster.com/dictionary/web%20crawler>
- SERCOP. (Diciembre 2014). Manual de usuario: Publicaciones de Emergencia [PDF]. *Sección 2. Fundamento Jurídico* (pp. 2-6). Extraído de <https://portal.compraspublicas.gob.ec/sercop/wp-content/uploads/downloads/2015/09/Publicaciones-de-emergencia-Entidades-contratantes.pdf>

ANEXO A: DOCUMENTACIÓN DEL CODIGO FUENTE

La documentación del código implementado para este trabajo de titulación se encuentra dentro del directorio `'docs/build/latex/'` bajo el nombre `dawscope.pdf`. Este documento fue autogenerado a partir de la descripción en código de todas las funciones y módulos. Este directorio es parte de la carpeta de proyecto homónima: `'DAWSCOPE.'`

ANEXO B: ESQUEMA GRÁFICO DE LA BASE DE DATOS

