

*Web scraping aplicado al proceso de clasificación y reconocimiento de gastos de empleados en
el caso de una empresa de la industria de la tecnología médica*

Informe Final de Pasantía Profesional

Fecha: 10/11/2022 - 10/06/2023

Estudiante:
Nicolas Julián Poveda Guacaneme

Profesores Tutores:
Luz Dary Montoya Ocampo
Carlos Orlando Rico Bonilla

Facultad de Contaduría Pública
Universidad Santo Tomás
Bogotá D.C., 2023

Tabla de Contenido

1. Resumen..... 3

2. Sinopsis técnica..... 4

3. Descripción general de la organización 7

4. Descripción de las actividades realizadas 8

5. Cumplimiento de objetivos 12

6. Impactos de la pasantía 12

7. Resultados obtenidos 14

8. Conclusiones 15

LISTA DE ANEXOS 17

Bibliografía: 27



1. Resumen

El presente documento tiene como objeto presentar las actividades y logros obtenidos en un proyecto de mejora técnico-contable llevado a cabo en, *Zimmer Biomet*, una compañía del sector de la industria de la tecnología médica, como parte del desarrollo de la pasantía profesional. La propuesta consiste en implementar las herramientas de web scraping para la automatización de los procesos de clasificación y reconocimiento de gastos, en este caso, para los reportados por los empleados pertenecientes al centro de costos de la filial de Zimmer Canadá. Este proyecto se sustentó en el diseño de códigos como VBA y Python para la extracción y el procesamiento de la información financiera.

Este trabajo es una respuesta a una necesidad que presentaba la empresa de optimizar sus procesos de clasificación y reconocimiento de gastos de empleados, los cuales se realizaban de manera manual extrayendo uno a uno los datos de cada empleado. Esto representaba una cantidad de tiempo considerable para la persona a cargo de la actividad. Se tuvo como objetivos el incrementar la precisión de la información, fortalecer la seguridad de los datos y aumentar la frecuencia del análisis de estos.

Se describirá detalladamente las actividades realizadas para la implementación del proyecto, incluyendo los requisitos clave de este, los campos necesarios para la extracción de datos, la búsqueda y selección de las herramientas de web scraping, el desarrollo del proceso y la aplicación y pruebas de este. Se analizan los problemas encontrados durante el trabajo, tales como la dificultad para acceder a toda la información en la página web objetivo, y se presentan soluciones para estas restricciones.

Finalmente, los resultados muestran un marcado aumento en la eficiencia del proceso de reconocimiento de gastos, lo que reduce en gran medida el tiempo y el esfuerzo que los empleados necesitan para hacerlo. Asimismo, se ha logrado una mayor precisión de la información registrada, reduciendo la posibilidad de errores. Este trabajo puede servir como base para futuros proyectos de automatización de procesos en la compañía y otras organizaciones similares.

2. Sinopsis técnica

Es importante mencionar que la gestión de los gastos de empleados resulta una tarea esencial para cualquier empresa, sin importar su tamaño o sector, esto se debe a que con una correcta gestión de estos se pueden realizar diferentes acciones de seguimiento, tales como provisionar el promedio de los gastos mes a mes, analizar los incrementos y/o decrementos de estos para su debido control y así evitar posibles errores o incluso fraudes.

Las nuevas tecnologías están desempeñando un papel cada vez más importante en el campo de la contabilidad, transformando la forma en que se procesan, reconocen, miden, presentan y analizan los datos financieros. En este sentido avances tecnológicos como *blockchain*, *data analytics*, *big data*, *artificial intelligence*, entre otras están revolucionando la práctica contable y ofreciendo beneficios significativos para las empresas y los profesionales de la contabilidad (Macías, et al., 2020; Vincet et al, 2020).

La automatización de tareas rutinarias y repetitivas está permitiendo a los contadores liberar tiempo y recursos para enfocarse en actividades de mayor valor añadido. Los sistemas de contabilidad automatizados y los programas de software especializados facilitan la captura y



procesamiento rápido de datos, lo que a su vez reduce errores y mejora la eficiencia en la gestión de la información financiera.

El uso de diversas herramientas y software que pueden ser utilizados para la gestión de gastos de empleados resulta vital, en el caso del presente trabajo la herramienta de *SAP concur solutions* es utilizada por la organización con el fin de que el empleado pueda reportar sus gastos realizados con la tarjeta corporativa y cargar los comprobantes de estos a la plataforma

En cuanto a las normas contables y fiscales, es importante mencionar que existen regulaciones específicas en relación con el reconocimiento y el control de los gastos de una empresa, por ejemplo, la *International Accounting Standard 19* de Beneficios a los Empleados. La implementación de un sistema de gestión de gastos que cumpla con estos estándares puede contribuir a mejorar la transparencia y la eficiencia en la gestión de los recursos financieros de la empresa.

En este sentido, es importante resaltar que la implementación de un sistema de gestión de gastos de empleados utilizando técnicas de web scraping puede generar importantes beneficios para la empresa, como la reducción de costos, la optimización del tiempo de trabajo, la prevención de fraudes y errores, y la mejora de la toma de decisiones financieras.

El web scraping es otra de las tecnologías de información emergente que se referenciaban previamente (Krotov & Tennyson, 2018), particularmente es una técnica dinámica utilizada para extraer datos no estructurados de sitios web y transformarlos en datos estructurados. También se identifica como extracción de datos web, scraping de datos web, recolección web o extracción de pantalla. El web scraping es una forma de minería de datos (Saurkar, Pathare y Gode, 2018). Este proceso se lleva a cabo a través de la utilización de software que pueda obtener la estructura de



las paginas web y esta sea analizada para su posterior uso, donde se extraerá los datos más importantes para el usuario y se podrá organizar y almacenar esta información.

La implementación de un sistema de web scraping para la gestión de gastos empresariales puede ofrecer múltiples beneficios, como la reducción de tiempos y costos en el proceso de registro de gastos, la mejora en la precisión de los datos registrados y la automatización de tareas repetitivas. El uso de estas herramientas puede ayudar a desarrollar un mejor análisis de la información y liberar a la persona de actividades que no agreguen mucho valor, tal y como menciona (Brightman, 2021), el RPA (*Robotic Process Automation*) refuerza la Contabilidad completando tareas no cualificadas. Fomenta la mejora continua al permitir a los contables utilizar sus conocimientos para mejorar continuamente la capacidad de los robots, al tiempo que se liberan para ser más analíticos y tener más visión de futuro.

Los procesos de automatización no solo pueden afectar de manera positiva a la empresa en cuanto a tiempos y prevención de errores, también puede ser de ayuda en temas relacionados con la auditoría o el *due diligence*, por ejemplo. Las empresas pueden utilizar controles automatizados, normalmente integrados en sus aplicaciones informáticas, de forma muy similar y con los mismos objetivos finales (Berman & Rubin, 2021) estos procesos permiten la mejora continua y el fortalecimiento de controles dentro de la empresa, desde un punto de vista tecnológico, el FoC (Futuro de los Controles) combina la automatización y las tecnologías avanzadas, como la automatización robótica de procesos (RPA), la inteligencia artificial (IA), el aprendizaje automático (ML) y el análisis predictivo, para agilizar y mejorar los controles, ofrecer inteligencia procesable y ayudar a las empresas a escalar más rápido y de mejor manera (Berman & Rubin, 2021).

3. Descripción general de la organización

Zimmer Biomet es una de las compañías más grandes y reconocidas en el campo de la tecnología médica. La empresa opera en más de 100 países y cuenta con más de 19,000 empleados en todo el mundo. Además, es conocida por su enfoque en la sostenibilidad y la responsabilidad social corporativa.

En cuanto a sus productos, Zimmer Biomet ofrece una amplia gama de dispositivos ortopédicos, desde prótesis de cadera y rodilla hasta equipos y herramientas quirúrgicas especializadas. La compañía ha desarrollado tecnologías innovadoras, como los sistemas de navegación quirúrgica, robots de asistencia en procesos quirúrgicos *ROSA (Robotic Surgical Assistant)* y los implantes personalizados, que permiten un mejor resultado en los procedimientos quirúrgicos y un proceso de recuperación más rápido y menos doloroso para los pacientes.

La investigación y el desarrollo son una parte integral de la cultura de la organización. La empresa invierte fuertemente en la innovación y trabaja en colaboración con médicos y expertos en la materia para desarrollar nuevas tecnologías que mejoren la atención de los pacientes. También ha establecido centros de investigación y desarrollo en todo el mundo, lo que demuestra su compromiso con la mejora continua.

Como se mencionó antes, otra característica destacable es su compromiso con la sostenibilidad y la responsabilidad social corporativa. La empresa se ha fijado objetivos ambiciosos, por ejemplo, en la reducción de las emisiones de gases de efecto invernadero, la gestión de residuos y la promoción de prácticas comerciales éticas y responsables.

En resumen, Zimmer Biomet es una empresa líder en tecnología médica con una gran



cantidad de productos ortopédicos innovadores y un enfoque en la investigación y el desarrollo. Además, su compromiso con la sustentabilidad y la responsabilidad social lo convierte en un modelo a seguir en la industria de la tecnología médica.

4. Descripción de las actividades realizadas

En la ejecución de la pasantía profesional se establecieron y desarrollaron las siguientes actividades:

1. Investigación sobre las herramientas de Web scraping y su aplicación en el proceso de gastos:

Fue una fase preliminar donde se realizó una investigación exhaustiva sobre las herramientas existentes y cómo se pueden aplicar en el proceso requerido. En esta etapa se analizó las diferentes opciones disponibles tales como: *Jsoup*, *Beautiful Soup*, *Puppeteer*, *Cheerio*, etc. Se tomó en cuenta sus ventajas, desventajas y su funcionalidad teniendo en cuenta el tipo de página web que en este caso es de tipo dinámica. También se examinó cual lenguaje de programación podría ser el más adecuado para la realización de este proyecto.

Al final se optó por usar el lenguaje de programación Python para la ejecución del *script* y las librerías de *request* y *selenium* para la extracción de datos.

2. Revisar los términos y condiciones de uso y las políticas de privacidad de SAP Concur solutions:

Antes de utilizar la herramienta de SAP se revisaron sus términos y condiciones de uso y las políticas de privacidad para asegurarse de que se cumplieran con los requisitos de la empresa y que los datos de los empleados estarían protegidos.



Se investigaron las políticas y regulaciones relevantes que se aplican al manejo de datos personales y financieros. También se revisaron los términos y condiciones de uso de *SAP Concur solutions* para entender cómo se podía utilizar el software de manera efectiva y dentro de los límites de la ley.

Además, se examinó el tema de la privacidad de los empleados y se validó y aseguró que la extracción de datos no afectará este aspecto fundamental. Por lo tanto, es primordial comprender cómo se protegen los datos de los empleados y cómo se pueden emplear de manera segura para mejorar el proceso de gestión de gastos.

3. Diseño del proceso de extracción de datos y definición de los campos requeridos para el proceso de gastos:

Una vez se seleccionó la herramienta de Web scraping, se diseñó el proceso de extracción de datos y se definieron los campos requeridos para el proceso de gastos, teniendo en cuenta los parámetros requeridos que debían ser llenados, tales como formularios y filtros, además de eso también se tuvo en cuenta la posición de los campos donde se encontraba la información que se iba a extraer. En esta etapa se determinó qué datos se necesitaban y cómo se debían extraer.

4. Configuración de las herramientas de Web scraping seleccionadas y pruebas de integración con los sistemas existentes de la empresa:

Una vez definido el proceso de extracción de datos, se configuraron las herramientas seleccionadas y se realizaron pruebas de integración con los sistemas existentes de la empresa para asegurarse de que funcionaran correctamente juntos. Se optó por usar “*Chrome Driver*”, un controlador de navegador que permite las interacciones con Google Chrome a



través de scripts, esta herramienta se uso como un intermediario entre el script y el navegador.

También se utilizó un IDE (entorno de desarrollo integrado) portable y libre de uso, esto con el fin de no generar fallas de seguridad o requerimientos al área de IT por el tema de licencias. El propósito del IDE es la de tener una interfaz amigable para el desarrollo y ejecución del script.

5. Desarrollo de scripts y algoritmos necesarios para extraer y procesar los datos:

En esta fase se desarrolló los scripts y algoritmos necesarios para extraer y procesar los datos de acuerdo con el diseño previamente establecido. (véase anexo A - F)

Una vez desarrollados los scripts y algoritmos, se implementó el proyecto en el proceso de gastos de la empresa y se realizaron pruebas para asegurarse de que funcionaba correctamente.

6. Capacitación de los empleados en el uso del nuevo proceso de gastos automatizado:

Una vez implementado el nuevo proceso de gastos automatizado, fue necesario capacitar al empleado a cargo de la actividad para que pudiera utilizarlo correctamente y aprovechar al máximo sus beneficios.

7. Análisis de los resultados obtenidos y evaluación de la efectividad del proceso de Web scraping en el manejo y registro de gastos de empleados:



Después de implementar el proceso de Web scraping, se realizó un análisis de los resultados obtenidos y se evaluó la efectividad del proceso en el manejo y registro de gastos de empleados.

8. *Identificación de oportunidades de mejora y recomendaciones para optimizar el proceso de Web scraping y su integración con los sistemas existentes de la empresa:*

Finalmente, se identificaron oportunidades de mejora y se realizaron recomendaciones para optimizar el proceso de Web scraping y su integración con los sistemas existentes de la empresa para mejorar la eficiencia y eficacia del proceso de gestión de gastos de los empleados.

La primera es el cambio de paradigma de programación, el paradigma usado para este plan de mejora fue el de programación estructurada, sin embargo, un paradigma de programación orientado a objeto (POO) optimizaría el proceso en términos de presentación de código, esto con el fin de hacerlo mas entendible al momento de que cualquiera otra empleada requiera consultarlo.

La segunda mejora es la implementación de este proyecto en procesos similares dado que en países como México y costa rica también se utiliza los servicios de *Concur solutions* para el manejo de los gastos de empleados.

Por último, la implementación de otros tipos de tecnología, tales como la identificación de patrones en los PDF para extraer y comparar las cadenas de texto o la implementación de OCR (reconocimiento óptico de caracteres) para la lectura de los soportes cargados en formatos de imagen, esto con el fin de identificar de manera más ágil si existe algún tipo de impuesto que pueda ser deducible.

5. Cumplimiento de objetivos

- *Incrementar la eficiencia y la precisión del proceso de gastos de empleados*

El plan de mejora permitió la automatización de la extracción y registro de datos de forma rápida y precisa, evitando errores humanos en la entrada manual de información. Además, esto redujo el tiempo y los recursos dedicados al procesamiento de los datos, permitiendo a los empleados dedicar más tiempo a otras tareas importantes.

- *Fortalecer la seguridad y protección de la información financiera*

En cuanto a la seguridad y protección de la información financiera, la implementación del proyecto permitió la extracción de información de una forma más segura y protegida que la entrada manual de datos, ya que el proceso se llevó a cabo de manera automatizada, minimizando la posibilidad de errores humanos y aumentando la confidencialidad de la información financiera.

- *Aumentar la frecuencia y la calidad del análisis de gastos*

La implementación del proceso de web scraping permitió una mayor frecuencia en la extracción de información y una alta cantidad de datos disponibles para el análisis. Esto facilitó a la empresa realizar análisis más precisos y detallados de los gastos de los empleados, identificar tendencias y áreas de mejora para reducir costos y aumentar la eficiencia.

6. Impactos de la pasantía

Los impactos de la pasantía se enumeran y describen así:

1. *Eficiencia operativa:* Estos tipos de tecnologías permiten agilizar y simplificar el proceso de clasificación y registro de gastos, eliminando tareas manuales repetitivas. Esto conduce a una mejora significativa en la eficiencia operativa, ya que se reduce drásticamente el tiempo necesario para completar estas actividades. Los empleados encargados de esta tarea pueden dedicar su tiempo y esfuerzo a otras actividades de mayor valor añadido para la organización.
2. *Ahorro de costos:* La automatización de procesos manuales evita la necesidad de asignar recursos humanos a tareas tediosas y propensas a errores. Esto se traduce en un ahorro de costos significativo a largo plazo, ya que se reduce la carga de trabajo y se optimiza el uso del personal disponible. También es importante resaltar que al reducir el personal necesario para realizar estas actividades se reducen costos relacionados con la compra de licencias para el uso del software donde se encuentra almacenada la información.
3. *Mayor precisión y fiabilidad:* Se reduce la probabilidad de cometer errores humanos, como ingresar datos incorrectos o clasificar de manera inadecuada los gastos. Esto garantiza la precisión y fiabilidad de los registros, lo que a su vez mejora la calidad de la información financiera y ayuda en el proceso de deducción de gastos. La precisión mejorada también ayuda a identificar áreas donde se pueden realizar ajustes o mejoras en el control de gastos.
4. *Cumplimiento normativo:* Al automatizar el proceso de clasificación y reconocimiento de gastos, se establece un sistema más consistente y confiable para el cumplimiento de



normativas y políticas internas de la organización. La automatización garantiza que los gastos se registren de acuerdo con las políticas y los procedimientos establecidos, lo que ayuda a mantener la transparencia y cumplir con los requisitos legales aplicables.

5. *Análisis y toma de decisiones basados en datos:* Proporciona una fuente de datos más completa y precisa para el análisis financiero. Esto facilita la identificación de patrones de gasto, la generación de informes más detallados y la toma de decisiones informadas. Los datos extraídos mediante web scraping pueden utilizarse para identificar oportunidades de ahorro, optimizar los presupuestos y mejorar la planificación financiera.

7. Resultados obtenidos

Con la implementación de este plan de mejora se logró una reducción significativa de los tiempos empleados para la realización de esta actividad, dado que, a comparación de el proceso manual al cual se le dedicaba en promedio 2 horas al diarias durante todo el mes para un total de 40 horas al mes, ahora con el uso del software el proceso puede llegar a demorarse 4 horas aproximadamente al mes sin necesidad de intervención humana. Además de esto, el programa se ejecuta en segundo plano, lo que significa que el empleado puede realizar actividades diferentes mientras se lleva a cabo la actividad. En una muestra que se tomó (véase el anexo G) se puede ver claramente la diferencia entre el empleado y el software.

Dado que los datos son extraídos directamente desde la posición exacta donde se encuentra en la web esto mejora la precisión y calidad de los registros, la automatización a través del scraping permitió disminuir los errores de ingreso de datos y la clasificación incorrecta de los gastos. Esto también facilitó la generación de informes financieros más precisos y una toma de decisiones más confiable basada en datos precisos y confiables (véase anexo H).

La implementación de la automatización permitió obtener una visión más clara y detallada de los gastos de los empleados. La extracción de datos mediante web scraping proporcionó información precisa y oportuna sobre los gastos realizados por cada empleado y su categorización. Esto brindó una mayor visibilidad y control sobre los gastos en toda la organización, lo que facilitó la identificación de patrones de gasto, el seguimiento de presupuestos asignados y la detección de posibles irregularidades o abusos en el uso de los recursos de la empresa.

Los datos obtenidos a través de la automatización permitieron una gestión financiera más efectiva. La disponibilidad de información precisa y actualizada sobre los gastos de los empleados facilitó el análisis financiero, la identificación de áreas de mejora y la toma de decisiones informadas. La organización pudo identificar oportunidades de ahorro, optimizar presupuestos y ajustar políticas de gastos de acuerdo con los datos recopilados.

8. Conclusiones

El trabajo realizado en la pasantía de automatización de clasificación y registro de gastos de empleados mediante web scraping ha demostrado ser altamente beneficioso para la organización. La automatización a través del web scraping ha mejorado significativamente la eficiencia del proceso de clasificación y reconocimiento de gastos de empleados de la empresa. Se ha reducido considerablemente el tiempo requerido para completar estas tareas, lo que ha permitido una asignación más efectiva de los recursos internos.

La precisión y la calidad de los registros de gastos se han mejorado gracias a la automatización. Al eliminar los errores humanos asociados con el proceso manual, se han

obtenido datos más precisos y confiables, lo que ha facilitado la conciliación de cuentas y la generación de informes financieros más exactos.

También es importante resaltar todas las oportunidades de mejora para el proceso que se pudieron identificar durante la implementación de este. El constante mantenimiento y mejora del programa será clave para el éxito de este a través del tiempo.

LISTA DE ANEXOS

Anexo A. Flujograma parte 1 18

Anexo B. Flujograma parte 2 19

Anexo C. Flujograma parte 3 20

Anexo D. Flujograma parte 4 21

Anexo E. Flujograma parte 5 22

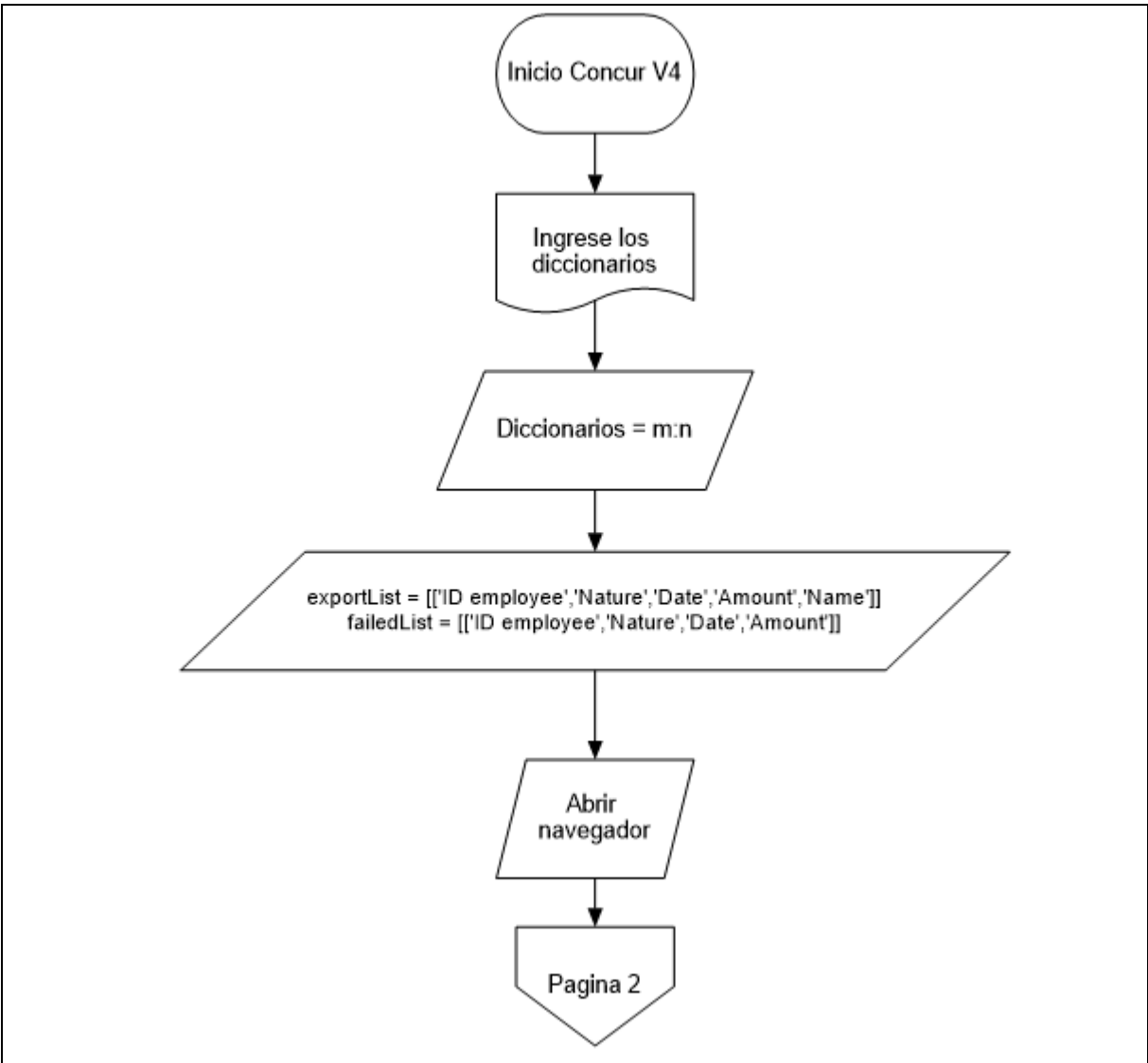
Anexo F. Flujograma parte 6 23

Anexo G. Muestra practicante vs software 24

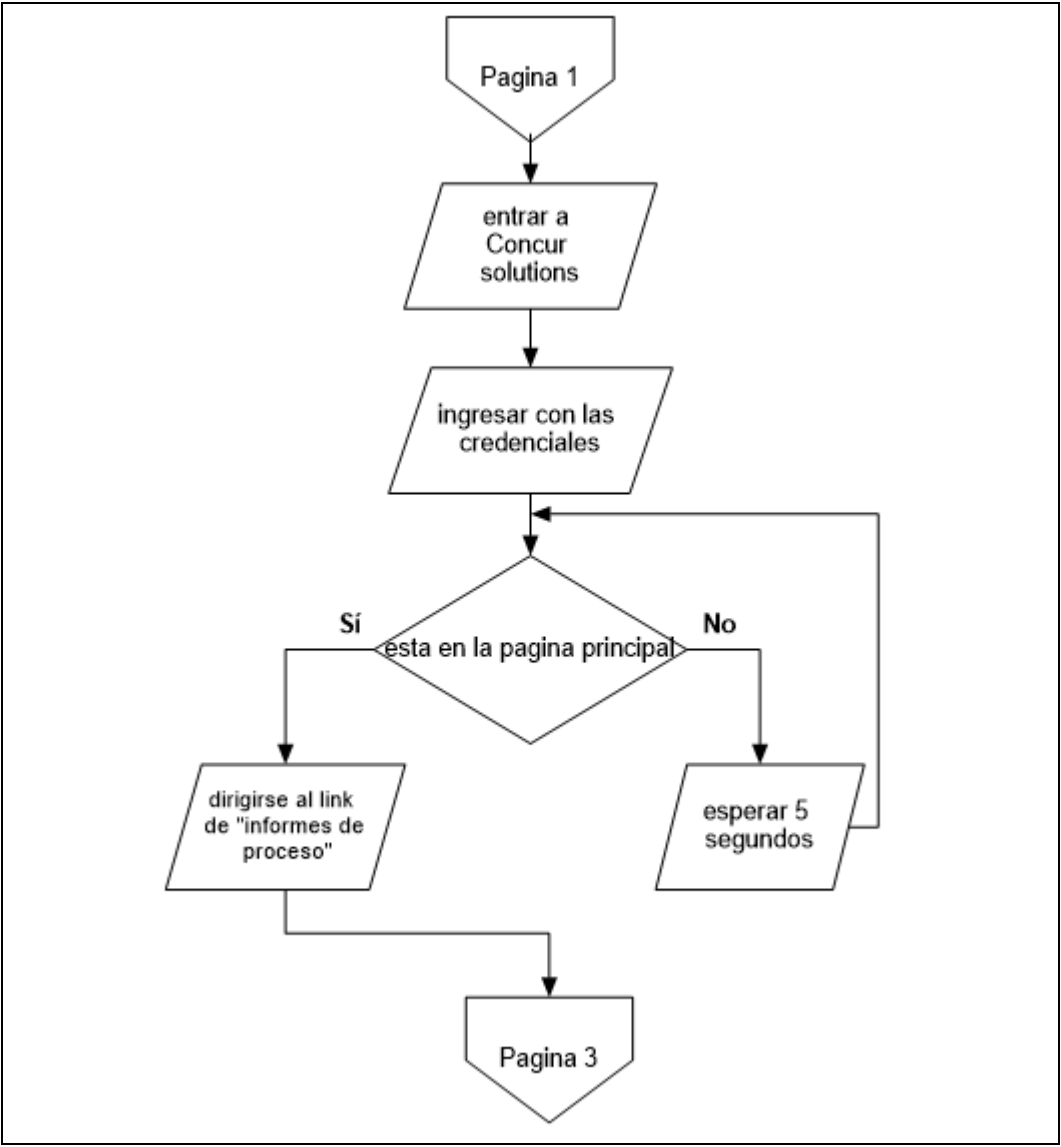
Anexo I. Resultados cualitativos 25

Anexo H. Reporte con los resultados 25

Anexo A. Flujograma parte 1

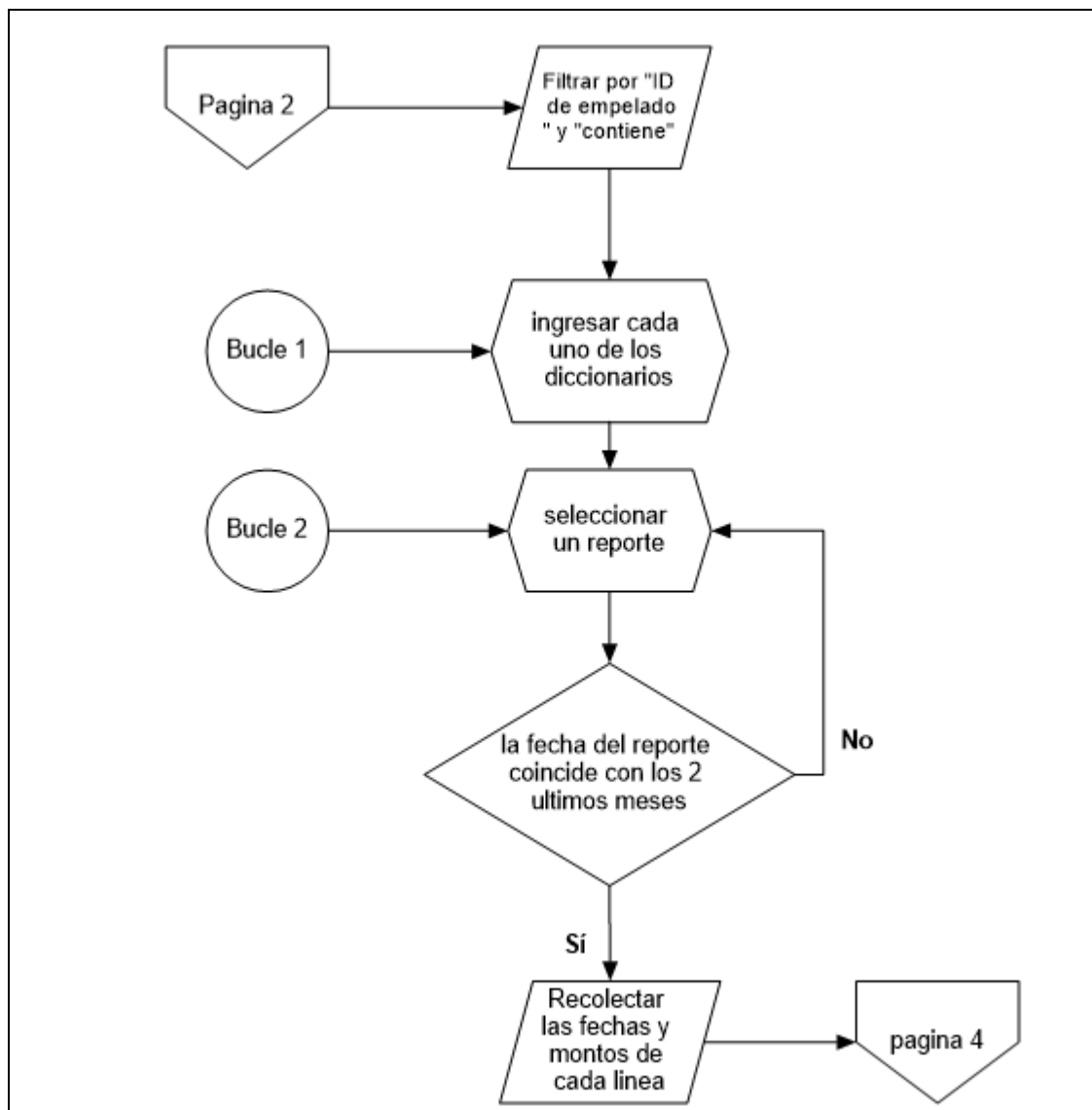


Anexo B. Flujograma parte 2



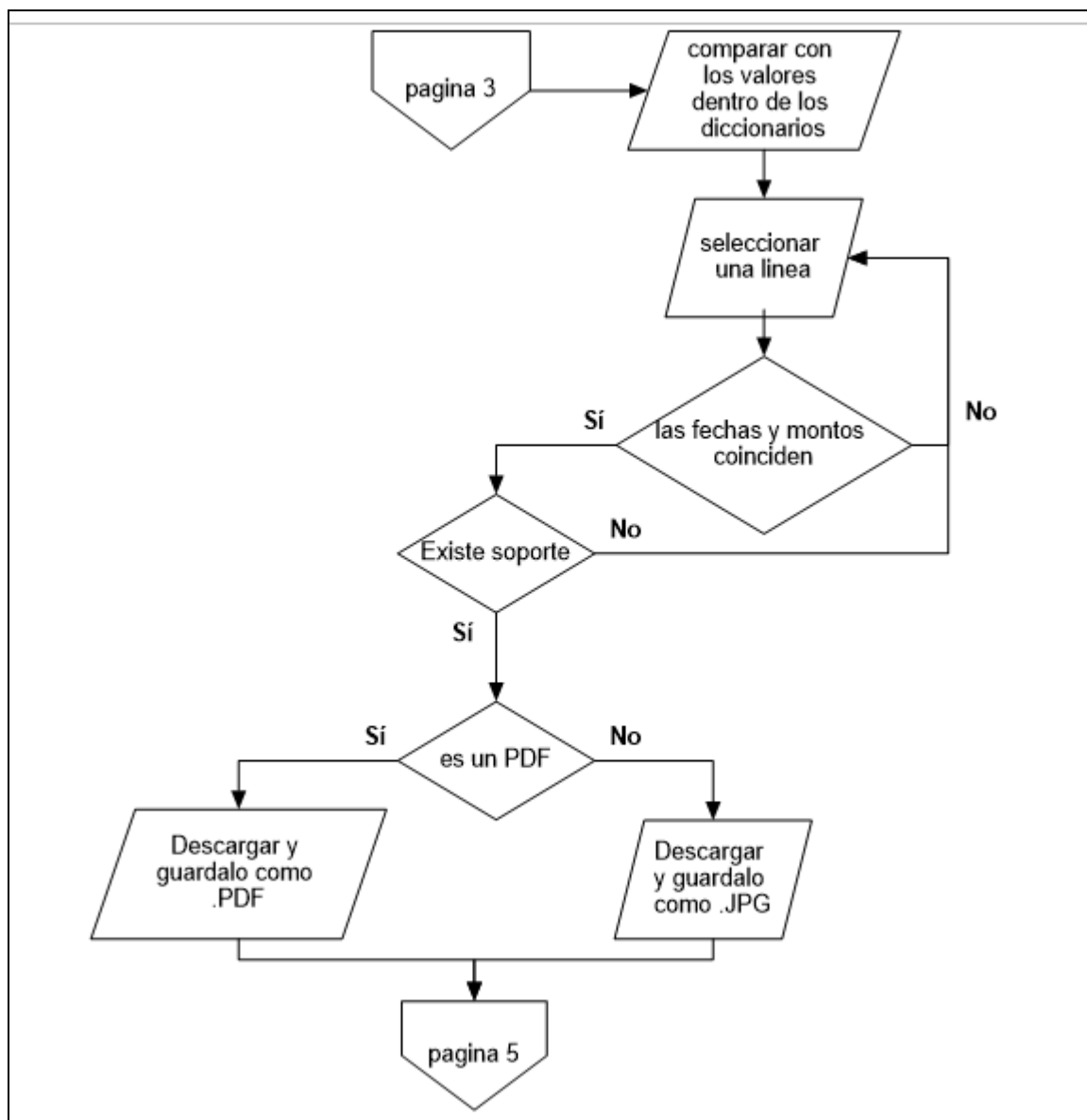


Anexo C. Flujograma parte 3

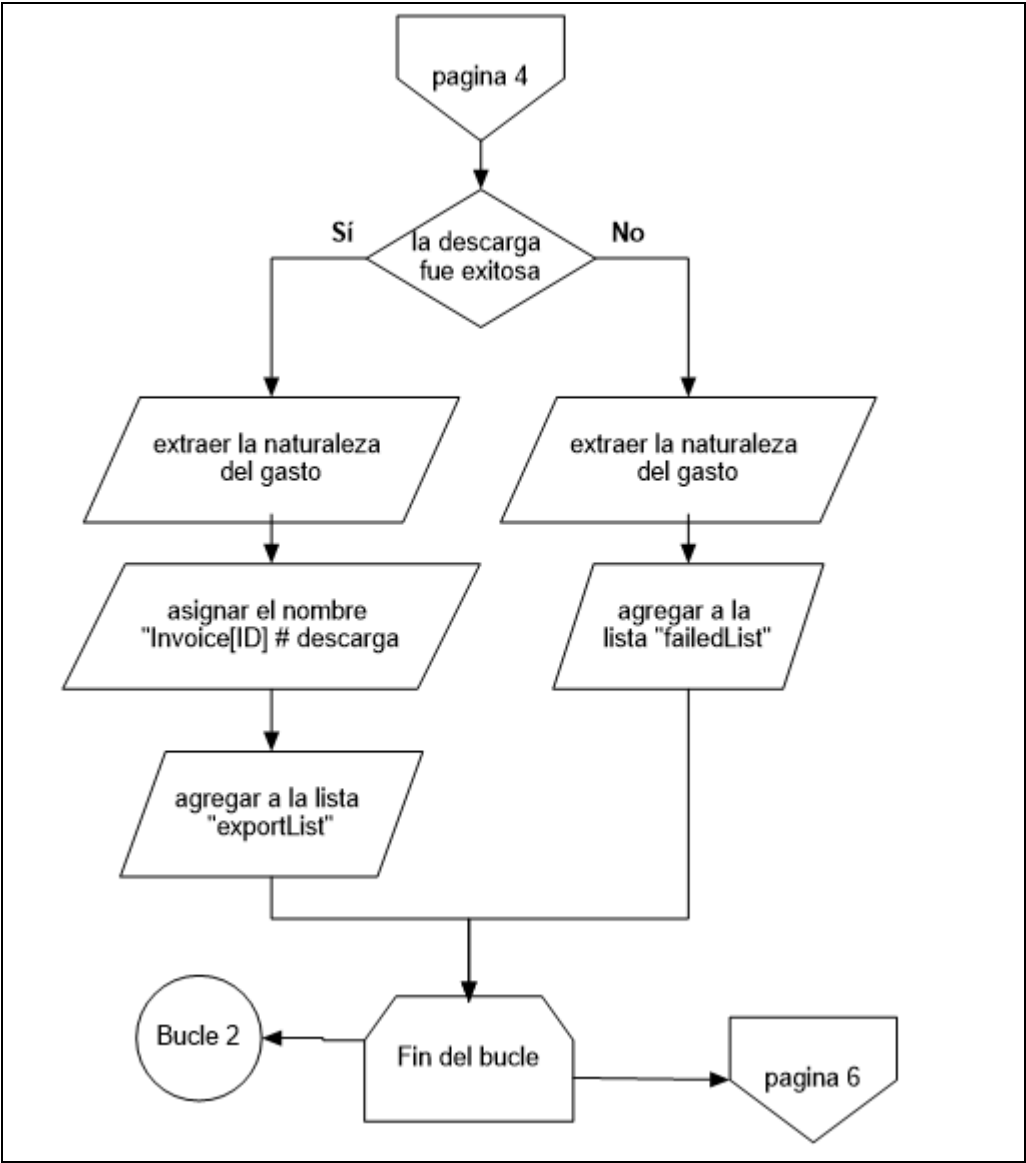




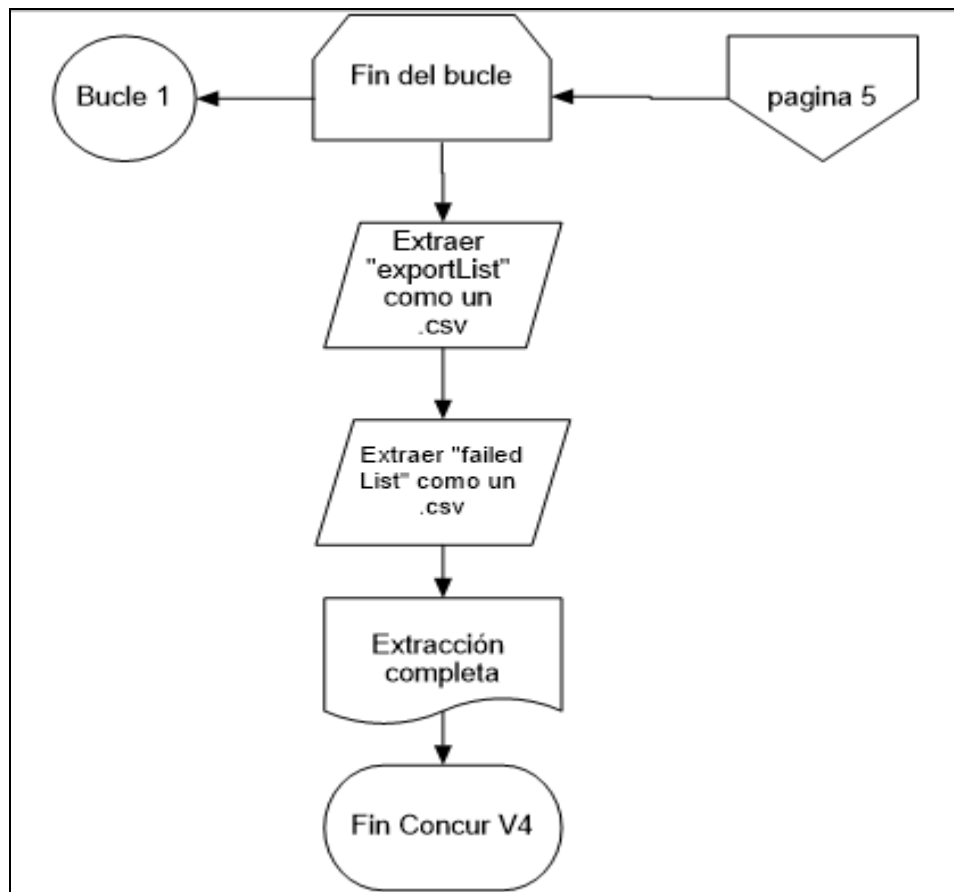
Anexo D. Flujograma parte 4



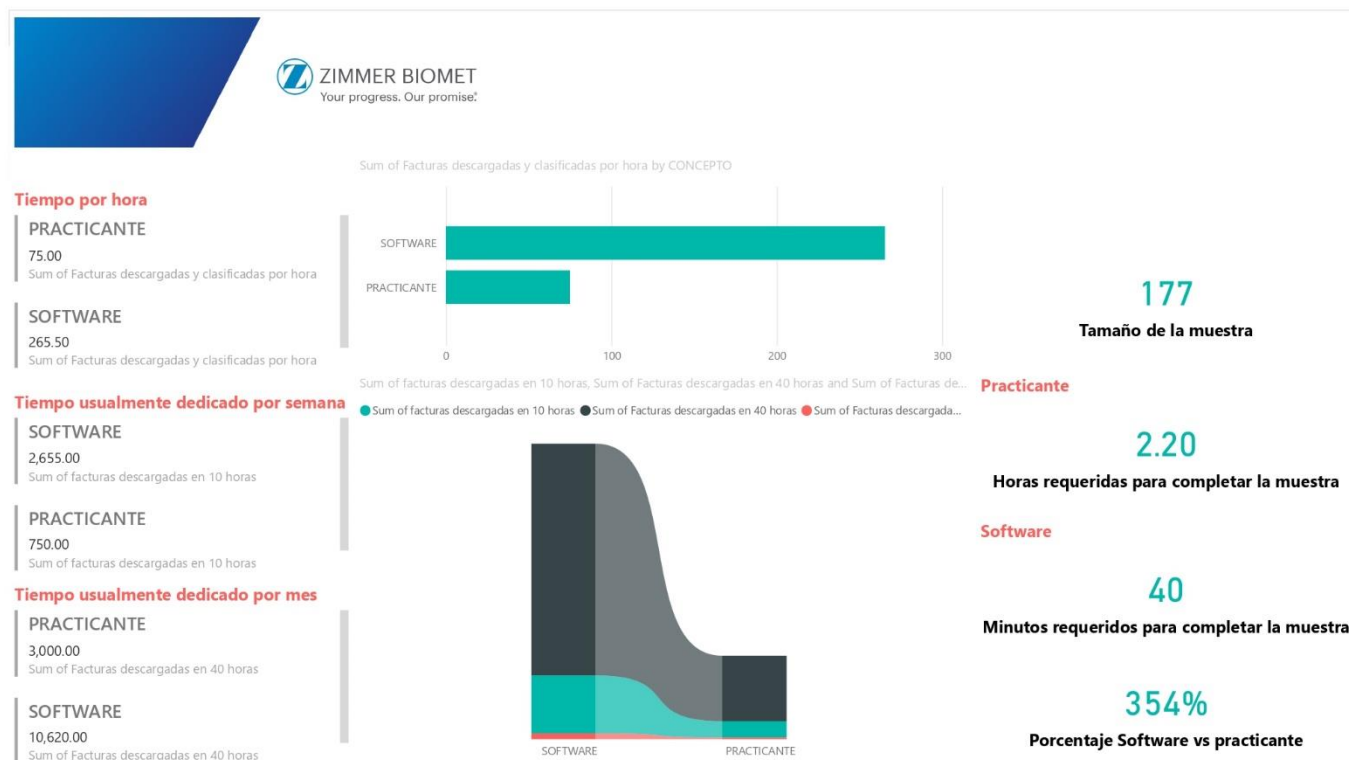
Anexo E. Flujograma parte 5



Anexo F. Flujograma parte 6



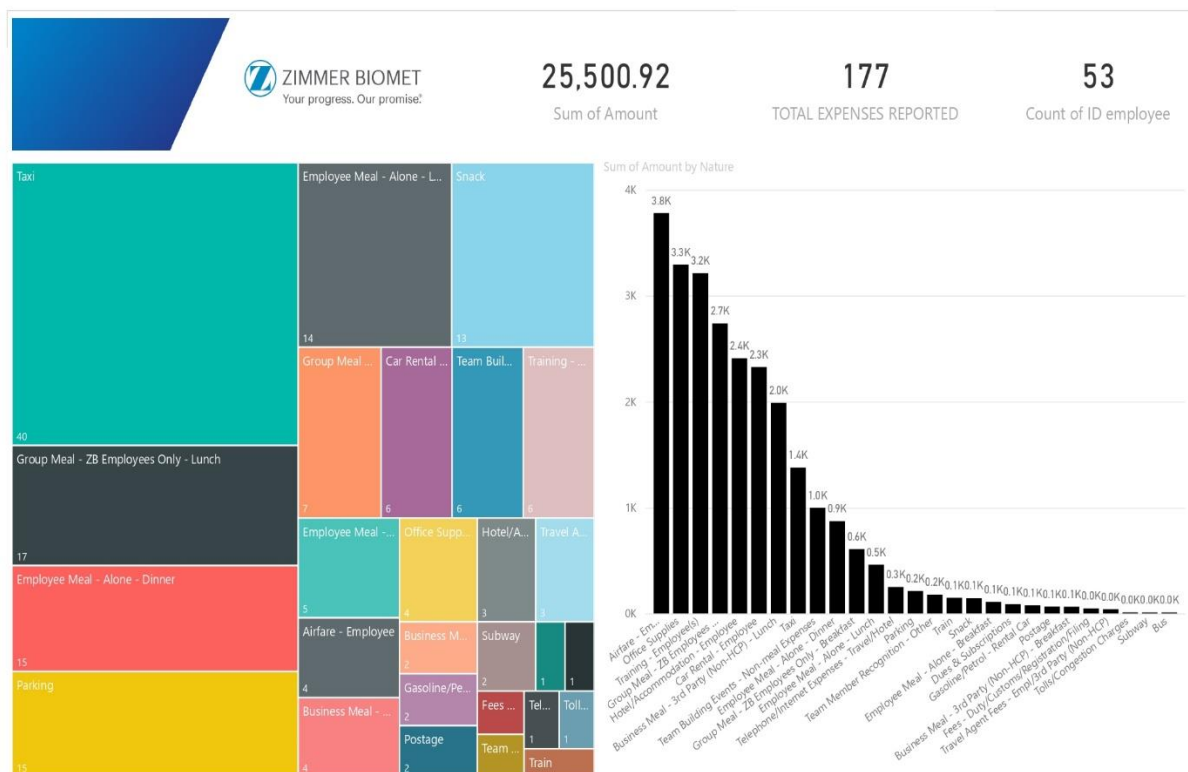
Anexo G. Muestra practicante vs software



Anexo H. Resultados cualitativos

SOFTWARE	PRACTICANTE
Desarrollo del proceso en segundo plano	Solo se puede realizar una actividad a la vez
Descarga ágil y precisa de facturas	Descarga manual de facturas y sujeta a errores
Requiere de una única persona para su funcionamiento.	Requiere de una o más personas para completar el reporte.
Requiere de aproximadamente 4 horas al mes para completar los reportes	Requiere de aproximadamente 40 horas al mes para completar los reportes

Anexo I. Reporte con los resultados



Bibliografía:

Berman, A., & Rubin, S. (06 de Diciembre de 2021). *The Future of Internal Controls: Embracing Advanced Automation*. Obtenido de deloitte: <https://www.deloitte.com/global/en/services/risk-advisory/blogs/the-future-of-internal-controls-embracing-advanced-automation.html>

Brightman, D. (06 de septiembre de 2021). *Blackline*. Obtenido de What Is Robotic Process Automation in Accounting?: <https://www.blackline.com/blog/robotic-process-automation-accounting/>

Krotov, V., & Tennyson, M. (2018). Nota de investigación: Scraping de datos financieros de la web utilizando el lenguaje R. *Journal of Emerging Technologies in Accounting*, 15(1), 169-181.

Macias, H. A., Farfán, M. A., & Rodríguez, B. A. (2020). Contabilidad digital: los retos del blockchain para académicos y profesionales. *Revista Activos*, 18(1).

Saurkar, A. V., Pathare, K. G., & Gode, S. A. (2018). *Revisión exhaustiva del web scraping*. International Journal on Future Revolution in Computer Science & Communication Engineering, 4(4). ISSN: 2454-4248.

Vincent, N. E., Igou, A., & Burns, M. B. (2020) Prepararse para los robots: Una propuesta de curso sobre automatización robótica de procesos. *Journal of Emerging Technologies in Accounting*, 17(2), 75-91.