



Facultad 2

Sistema Informático para la Vigilancia Tecnológica del Centro de Telemática de la Universidad de la Ciencias Informáticas.

Trabajo de Diploma para optar por el título de
Ingeniero en Ciencias Informáticas

Autores: Claudia Benítez Rodríguez
Néstor J. Pérez González

Tutores: Ing. Suany Leyva Hernández
Ing. Cyndi Lorenzo Jorge

Co-tutor: Ing. Bárbaro Guillermo Barroso Gómez

Consultante: Dr. Jorge Sergio Menéndez Pérez

La Habana

Junio 2018



"En tiempos de cambio, quienes estén abiertos al aprendizaje se adueñarán del futuro, mientras que aquellos que creen saberlo todo estarán bien equipados para un mundo que ya no existe."

Eric Hoffer

DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis que tiene por título: “Sistema informático para la Vigilancia Tecnológica del Centro de Telemática de la Universidad de la Ciencias Informáticas” y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo. Para que así conste firmamos la presente a los ____ días del mes de _____ del año 2018.

Claudia Benítez Rodríguez

Néstor J. Pérez González

Firma del Autor

Firma del Autor

Suany Leyva Hernández

Cyndi Lorenzo Jorge

Firma del Tutor

Firma del Tutor

DATOS DE CONTACTO

Autor:

Claudia Benítez Rodríguez

Universidad de las Ciencias Informáticas

Email: cbenitez@estudiantes.uci.cu

Autor:

Néstor J. Pérez González

Universidad de las Ciencias Informáticas

Email: njperez@estudiantes.uci.cu

Tutor:

Suany Leyva Hernández

Universidad de las Ciencias Informáticas

Email: slhernandez@uci.cu

Tutor:

Cyndi Lorenzo Jorge

Universidad de las Ciencias Informáticas

Email: cyndi@uci.cu

RESUMEN

La gestión estratégica de la información científico-tecnológica resulta cada vez más importante para innovar y sobrevivir en un entorno complejo y cambiante como el actual. El Centro de Telemática de la Universidad de las Ciencias Informáticas requiere de información actualizada de los productos y servicios de la competencia, para poder desarrollar sistemas competitivos en el mercado internacional. Reunir toda esa información de forma manual es complicado, no garantiza que esté completa y correcta, ni totalmente actualizada, por lo que se definió como objetivo de la investigación desarrollar una aplicación para el proceso de Mercadotecnia del Centro de Telemática que permita informatizar el proceso de vigilancia tecnológica de los productos y servicios del mismo.

Se determinó emplear la técnica de *web scraping* y recuperación de información usando lógica difusa para darle solución al problema. Se definió como metodología de desarrollo *Extreme Programming (XP)*, como tecnologías y herramientas de desarrollo, XILEMA como estrategia marcaria, el framework Django, los lenguajes Python, HTML, JavaScript y CSS, el IDE JetBrains-PyCharm, las bibliotecas Beautiful Soup y jQuery, el sistema gestor de bases de datos PostgreSQL y el kit de herramientas Bootstrap.

Se obtuvo una aplicación que realiza vigilancia tecnológica a través de las técnicas empleadas permitiendo la realización de la búsqueda de información en la web, clasificándose a su vez según el peso de importancia para los usuarios, conformando un documento en formato PDF con la información obtenida y enviándose este por correo.

Palabras claves: búsqueda, recopilación, utilización, contenido web, *web scraping*.

Índice

INTRODUCCIÓN	1
---------------------------	---

CAPÍTULO 1: FUNDAMENTOS DE LA INVESTIGACIÓN, METODOLOGÍA Y TECNOLOGÍAS EMPLEADAS	1
---	---

1.1 Marco conceptual.....	1
1.2 Consumo y reutilización de contenido web. Estado del arte	8
Herramientas a nivel internacional.....	8
1.2.1 Import.io	8
1.2.2 Helium Scraper.....	8
1.2.3 Web Data Extractor.....	9
1.2.4 OutWit Hub Review.....	9
1.2.5 WebHarvy	9
1.2.6 Octoparse	10
1.2.7 SoftVT	11
1.2.8 Vigiale.....	11
1.2.9 Hontza.....	11
1.2.10 Web Scrap.....	12
1.3 Metodología.....	12
1.4 Tecnologías y herramientas de desarrollo	15
1.4.1 Python 2.7.14	15
1.4.2 Django 1.11.6.....	16
1.4.3 JetBrains-PyCharm 2017 2.2.....	16
1.4.4 BeautifulSoup 4.6.....	16
1.4.5 Requests 2.7.0	17
1.4.6 HTML5	17
1.4.7 CCS3.....	18
1.4.8 JavaScript (JS).....	18
1.4.9 jQuery 1.9.11	18
1.4.10 Bootstrap 3	18
1.4.11 RSS.....	18
1.4.12 PostgreSQL 9.4.....	19
1.4.13 Visual Paradigm 15.0 Community Edition	19
1.5 Conclusiones del capítulo	19

CAPÍTULO 2: CARACTERÍSTICAS DEL SISTEMA. EXPLORACIÓN, PLANIFICACIÓN Y DISEÑO	21
---	----

2.1 Características de la propuesta del sistema	21
2.2 Lista de reservas de producto	22
2.3 Características no funcionales del sistema	23
2.4 Exploración.....	24
2.4.1 Historia de usuario (HU).....	24

2.5 Planeación	33
2.5.1 Estimación de esfuerzo por HU	33
2.5.2 Plan de iteraciones	34
2.5.3 Plan de duración de las iteraciones	34
2.5.4 Plan de entrega	35
2.6 Diseño	35
2.6.1 Patrón y estilo de arquitectura	36
2.6.2 Patrones de diseño	37
2.6.3 Tarjetas CRC	41
2.7 Diseño de la Base de Datos	43
2.8 Algoritmo a desarrollar	44
2.9 Conclusiones del capítulo	46
CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBAS.	48
3.1 Estándares de codificación	48
3.2 Iteraciones	49
3.2.1 Pruebas	49
3.3 Iteración 1	50
3.4 Iteración 2	52
3.5 Iteración 3	54
3.6 Iteración 4	57
3.7 Iteración 5	61
3.8 Comprobación del aprendizaje del algoritmo	64
3.9 Conclusiones del capítulo	66
CONCLUSIONES	67
RECOMENDACIONES	68
REFERENCIAS BIBLIOGRÁFICAS	69
BIBLIOGRAFÍA.....	73
ANEXOS. CASOS DE PRUEBA DE ACEPTACIÓN.....	77

ÍNDICE DE TABLAS

Tabla 1. Criterios para la selección apropiada de la metodología	12
Tabla 2. Parámetros para las HU	25
Tabla 3. HU – Autenticar usuario.....	25
Tabla 4. HU - Gestionar temáticas.	26
Tabla 5. HU - Gestionar de selectores.....	27
Tabla 6. HU - Gestionar URLs.....	27
Tabla 7. HU - Iniciar y detener el monitoreo de las URLs.....	28
Tabla 8. HU - Configurar automáticamente el monitoreo.....	29
Tabla 9. HU - Listar el registro de la búsqueda y el número de URLs analizadas.....	29
Tabla 10. HU - Gestionar los resultados de las búsquedas.....	30
Tabla 11. HU - Generar documento PDF con resultados seleccionados.....	31
Tabla 12. HU - Configurar envío de informes por correo.....	31
Tabla 13. HU – Configurar envío de forma automática de los resultados.....	32
Tabla 14. Estimación de esfuerzo por HU.....	33
Tabla 15. Plan de duración de las iteraciones.....	34
Tabla 16. Plan de entrega.....	35
Tabla 17. Tarjeta CRC #1.....	41
Tabla 18. Tarjeta CRC #2.....	41
Tabla 19. Tarjeta CRC #3.....	41
Tabla 20. Tarjeta CRC #4.....	42
Tabla 21. Tarjeta CRC #5.....	42
Tabla 22. Tarjeta CRC # 6	42
Tabla 23. Tarjeta CRC # 7	43
Tabla 24. Tarjeta CRC # 8	43
Tabla 25. Tarea de Ingeniería - Iniciar el monitoreo de las URL.....	50
Tabla 26. Tarea de Ingeniería - Detener el monitoreo de las URL.....	50
Tabla 27. Caso de Aceptación - Iniciar y detener el monitoreo de las URL - Datos Válidos.....	51
Tabla 28. Caso de Aceptación - Iniciar y detener el monitoreo de las URL - Datos Inválidos.....	52
Tabla 29. Tarea de Ingeniería - Añadir los resultados de las búsquedas.....	52
Tabla 30. Tarea de Ingeniería - Listar los resultados de las búsquedas.....	53
Tabla 31. Tarea de Ingeniería - Mostrar los resultados de las búsquedas.....	53
Tabla 32. Tarea de Ingeniería - Eliminar los resultados de las búsquedas.....	53
Tabla 33. Tarea de Ingeniería – Modificar temáticas.....	54
Tabla 34. Tarea de Ingeniería - Añadir temáticas.....	54
Tabla 35. Tarea de Ingeniería - Mostrar temáticas.....	55
Tabla 36. Tarea de Ingeniería - Modificar clasificación.....	55
Tabla 37. Tarea de Ingeniería - Eliminar temáticas.....	55
Tabla 38. Tarea de Ingeniería - Configurar automáticamente el monitoreo.....	56
Tabla 39. Tarea de Ingeniería - Añadir selectores.....	57
Tabla 40. Tarea de Ingeniería - Modificar selectores.....	57
Tabla 41. Tarea de Ingeniería - Mostrar selectores.....	58
Tabla 42. Tarea de Ingeniería - Listar selectores.....	58
Tabla 43. Tarea de Ingeniería - Eliminar selectores.....	58
Tabla 44. Tarea de Ingeniería - Añadir URL.....	59

Tabla 45. Tarea de Ingeniería - Modificar URL.....	59
Tabla 46. Tarea de Ingeniería - Mostrar URL.....	59
Tabla 47. Tarea de Ingeniería - Listar URL.....	60
Tabla 48. Tarea de Ingeniería - Eliminar URL.....	60
Tabla 49. Tarea de Ingeniería - Listar el registro de la búsqueda y el número de URLs analizadas.	60
Tabla 50. Tarea de Ingeniería - Generar documento PDF con los resultados seleccionados.....	61
Tabla 51. Tarea de Ingeniería - Autenticarse usuario.....	62
Tabla 52. Tarea de Ingeniería - Gestionar envío de informes por correo.....	62
Tabla 53. Tarea de Ingeniería - Envío de forma automática de los resultados.....	62
Tabla 54. Primer resultado para la búsqueda de la temática Call Center.....	65
Tabla 55. Segundo resultado para la búsqueda de la temática Call Center.....	65
Tabla 56. Caso de Prueba de Aceptación - Autenticar Usuario - Datos Válidos.....	77
Tabla 57. Caso de Prueba de Aceptación - Autenticarse Usuario - Datos Inválidos.....	77
Tabla 58. Caso de Prueba de Aceptación - Añadir temáticas - Datos Válidos.....	78
Tabla 59. Caso de Prueba de Aceptación - Añadir temáticas - Datos Inválidos.....	78
Tabla 60. Caso de Prueba de Aceptación - Mostrar temáticas - Datos Válidos.....	79
Tabla 61. Caso de Prueba de Aceptación - Listar temáticas - Datos Válidos.....	79
Tabla 62. Caso de Prueba de Aceptación - Modificar temáticas - Datos Válidos.....	80
Tabla 63. Caso de Prueba de Aceptación - Modificar temáticas - Datos Inválidos.....	80
Tabla 64. Caso de Prueba de Aceptación - Eliminar temáticas - Datos Válidos.....	81
Tabla 65. Caso de Prueba de Aceptación - Añadir Selectores - Datos Válidos.....	81
Tabla 66. Caso de Prueba de Aceptación - Listar Selectores - Datos Válidos.....	82
Tabla 67. Caso de Prueba de Aceptación - Modificar Selectores - Datos Válidos.....	82
Tabla 68. Caso de Prueba de Aceptación - Modificar Selectores - Datos Inválidos.....	83
Tabla 69. Caso de Prueba de Aceptación - Eliminar Selectores - Datos Válidos.....	83
Tabla 70. Caso de Prueba de Aceptación - Añadir URL - Datos Válidos.....	84
Tabla 71. Caso de Prueba de Aceptación - Añadir URL - Datos Inválidos.....	84
Tabla 72. Caso de Prueba de Aceptación - Añadir URL según los resultados de internet - Datos Válidos.....	86
Tabla 73. Caso de Prueba de Aceptación - Listar URL - Datos Válidos.....	86
Tabla 74. Caso de Prueba de Aceptación - Modificar URL - Datos Válidos.....	87
Tabla 75. Caso de Prueba de Aceptación - Modificar URL - Datos Inválidos.....	87
Tabla 76. Caso de Prueba de Aceptación - Eliminar URL - Datos Válidos.....	88
Tabla 77. Caso de Aceptación - Configurar automáticamente el monitoreo - Datos Válidos.....	88
Tabla 78. Caso de Aceptación - Configurar automáticamente el monitoreo - Datos Inválidos.....	89
Tabla 79. Caso de Aceptación - Mostrar trazas de las búsquedas realizadas - Datos Válidos.....	89
Tabla 80. Caso de Aceptación - Listar resultados de las búsquedas - Datos Válidos.....	90
Tabla 81. Caso de Aceptación - Eliminar resultados de las búsquedas - Datos Válidos.....	90
Tabla 82. Caso de Aceptación - Modificar resultados de las búsquedas - Datos Válidos.....	91
Tabla 83. Caso de Aceptación - Generar documento PDF con resultados seleccionados - Datos Válidos.....	91
Tabla 84. Caso de Aceptación - Configurar envío de informes por correo - Datos Válidos.....	92
Tabla 85. Caso de Aceptación - Configurar envío de informes por correo - Datos Inválidos.....	93

ÍNDICE DE FIGURAS

Fig. 1 Función de SRI difuso.....	5
Fig. 2 Resultado de la función para cada bloque de información.	5
Fig. 3 Resultados del empleo de los criterios.....	14
Figura 4. Propuesta del sistema.	22
Fig. 5 Diagrama del patrón Modelo Plantilla Vista (MTV)	36
Fig. 6 Evidencia del uso del patrón Alta cohesión	37
Fig. 7 Evidencia del uso del patrón Bajo acoplamiento.....	38
Figura 8. Evidencia del uso del patrón Creador	38
Fig. 9 Evidencia del uso del patrón Experto	39
Figura 10. Evidencia del uso del patrón Bridge	40
Figura 11. Evidencia del uso del patrón Singleton.....	40
Fig. 12 Modelo físico de la base de datos	44
Fig. 13 Resultados de las pruebas unitarias de la primera iteración	51
Fig. 14 Resultados de las pruebas unitarias de la segunda iteración.....	54
Fig. 15 Resultados de las pruebas unitarias de la tercera iteración	57
Fig. 16 Resultados de las pruebas unitarias de la cuarta iteración.....	61
Fig. 17 Resultados de las pruebas unitarias de la quinta iteración.....	63
Fig. 18 Resultados de los Casos de Pruebas de Aceptación.....	64

INTRODUCCIÓN

La gestión estratégica de la información científico-tecnológica resulta importante para innovar y sobrevivir en un entorno complejo y cambiante como el actual. Los mercados se hacen globales, el riesgo tecnológico requiere ser gestionado y la innovación abierta¹ emerge para transformar los modelos de negocio de las organizaciones. Además, la Internet permite acceder a una gran cantidad de información que, muchas veces, puede resultar inabarcable despertando el riesgo de la sobreinformación o infoxicación².

Empresas e investigadores comienzan a asumir la gestión del conocimiento y la tecnología como parte vital de sus estrategias de innovación y negocio, bien para alcanzar la excelencia científica, participar en proyectos de investigación colaborativa, incrementar la capacidad de adopción tecnológica de la organización o apostar por la internacionalización.

Entre las fases de los procesos de la gestión del conocimiento y la tecnología se encuentra la Vigilancia Tecnológica (VT) e Inteligencia Competitiva. (OVTT) Según la norma UNE 166006:2006 Ex Gestión de la I+D+i³ (Sistema de Vigilancia Tecnológica), se define que: “La Vigilancia Tecnológica es un proceso organizado, selectivo y permanente, de captar información del exterior y de la propia organización sobre ciencia y tecnología, seleccionarla, analizarla, difundirla y comunicarla, para convertirla en conocimiento para tomar decisiones con menor riesgo y poder anticiparse a los cambios”. (La vigilancia tecnológica en la gestión de proyectos de I+D+i: recursos y herramientas., 2006)

Al emplearla otorga ciertas ventajas a las empresas como:

- Información de valor a las empresas y organizaciones.
- Incremento de las posibilidades de generar acciones que permitan anticiparse al mercado.
- Reducción de riesgos asociados a las actividades de I+D+i.
- Fomenta la cooperación, abren nuevas oportunidades de mercado.
- Posibilita la generación de nuevos proyectos de I+D+i.
- Permite identificar señales de cambio (emergentes) que pueden llegar a tener gran impacto para la organización.

En Cuba, aunque es incipiente la aplicación de la VT, se muestran resultados palpables en el sector biotecnológico. La industria del software también precisa de una gestión eficiente de la información

¹ nueva estrategia de innovación mediante la cual las empresas van más allá de sus límites y desarrollan la cooperación con organizaciones o profesionales externos

² Infoxicación: Intoxicación por sobreinformación.

³ Investigación, desarrollo e innovación.

tecnológica y científica, por lo que es necesario su aplicación en los diferentes centros productores de software del país.

La Universidad de las Ciencias Informáticas (UCI) es una institución cubana creada en 2002 que se caracteriza por su modelo de formación novedoso que vincula la docencia y la producción de software. El desarrollo de producción de software de la UCI está distribuido en 14 centros y orientado a cinco áreas temáticas de alto impacto: Salud, Educación, Telemática, Empresa e Industria y Administración Pública. Los centros trabajan sobre estas áreas, delimitándolas en líneas más específicas. (Dirección de Comunicación Institucional 2016)

Uno de estos centros es el de Telemática (TLM), que desarrolla productos y servicios (P/S) relacionados con las Telecomunicaciones y la Seguridad Informática. En TLM existen varios procesos, uno de estos es el proceso de Mercadotecnia, que tiene la responsabilidad de reunir toda la información necesaria para elaborar los expedientes comerciales de los P/S. Cada expediente recoge elementos comerciales y técnicos; la información comercial incluye el modelo de negocio, las principales funcionalidades de los productos de la competencia, los precios de referencia del mercado y las estrategias de precios, de ahí la necesidad de tener una vigilancia constante sobre los sistemas similares a los desarrollados por el centro.

El proceso de realización del informe comienza cuando el Asesor de Mercadotecnia identifica en Internet las fuentes de información relacionadas con los P/S; después valida, recopila, clasifica, organiza y analiza todos los datos requeridos para así conformar el documento; este proceso es necesario que se haga lo más adecuado posible debido a que a través de los datos obtenidos de la competencia se puede mejorar el desarrollo de los productos creados en el centro, para así poder colocarlos en el mercado mundial. La búsqueda de la información lleva un consumo excesivo de tiempo, pues implica examinar la mayor cantidad de páginas posibles, y extraer los datos para después clasificarlos y analizarlos. Existen varios factores en contra del desarrollo exitoso del informe, como la ocurrencia del error humano al cambiar datos o malinterpretarlos, la pérdida de la información al no guardarse correctamente o ante la rotura del medio de almacenamiento; además, se encuentra el riesgo de desactualización de la información, teniendo en cuenta que el avance de las tecnologías es constante, y la publicación de resultados también. A raíz de lo planteado anteriormente, se puede afirmar que la obtención manual todos los elementos requeridos para una exitosa toma de decisiones es altamente compleja.

Teniendo en cuenta la situación problemática existente, los autores de la tesis han identificado el siguiente **Problema de la investigación:** ¿Cómo informatizar el proceso de vigilancia tecnológica de los productos que desarrolla el Centro de Telemática de la UCI?

En función de resolver el problema anterior se definen como **objeto de estudio**: el proceso de vigilancia tecnológica, centrándose en el **campo de acción** el proceso de vigilancia tecnológica de los productos que desarrolla el Centro de Telemática. Además, se propone como **objetivo general** desarrollar una aplicación para el proceso de Mercadotecnia del Centro de Telemática, que permita informatizar el proceso de vigilancia tecnológica de los productos que desarrolla el centro.

Para darle solución a la situación problemática se formulan las siguientes **preguntas científicas**:

- ¿Cuáles son las tecnologías que permiten realizar la VT?
- ¿Qué procesos intervienen en la VT?
- ¿Qué herramientas y tecnologías son necesarias a emplear durante el desarrollo de la aplicación?
- ¿Cuál es el plan y diseño del sistema a desarrollar?
- ¿Qué resultados tendrá la aplicación informática a desarrollar?

Para dar cumplimiento al objetivo planteado, se trazan las siguientes **tareas a desarrollar** en la investigación:

- Caracterización y análisis de tecnologías que permitan realizar VT.
- Análisis de los procesos que intervienen en la VT.
- Selección de herramientas y tecnologías a emplear para desarrollar la solución al problema de la investigación.
- Planeación y diseño del sistema a desarrollar.
- Realización de pruebas para la comprobación de la correcta utilización del sistema.

Para facilitar el cumplimiento del objetivo propuesto y de las tareas de investigación se emplean los métodos teóricos y empíricos siguientes:

Métodos teóricos:

- **Analítico–Sintético:** Se empleó para analizar materiales bibliográficos y por consiguiente conformar el estado del arte, en el que se definieron conceptos relacionados con la investigación. Permitted sintetizar dentro del conjunto de herramientas para funcionalidades similares y diferentes de aplicaciones web que hacen *web scraping*⁴ y recuperación de información.

⁴ Español: Raspado web

- **Histórico-Lógico:** Este método permite estudiar todo lo referente a la vigilancia tecnológica, para así obtener un conocimiento histórico de su desarrollo y comportamiento a nivel internacional como nacional.
- **Modelación:** Este método posibilitó la representación de las propiedades y funcionalidades del sistema a desarrollar, detallando su efectividad para resolver el problema de investigación planteado, utilizando artefactos como guía durante el diseño de la aplicación.

Métodos empíricos:

- ✓ **Observación:** Se utilizó para obtener el comportamiento real del problema planteado.
- **Entrevista:** Se consultó a personas que tienen conocimientos sobre los procesos de vigilancia tecnológica, así como de los métodos de recuperación de la información y la lógica difusa.
- **Medición:** La aplicación obtenida como resultado de la investigación será examinada a través de las pruebas unitarias y de aceptación comprobando su correcto funcionamiento.

Con el objetivo de estructurar y organizar el desarrollo del trabajo investigativo de una forma coherente, se divide en 3 capítulos:

Capítulo 1: Fundamentos de la investigación, metodología y tecnologías empleadas: Se abordan los elementos teóricos más importantes que sirven de base para la realización de la investigación; también se proporciona una descripción de las herramientas y metodología de desarrollo de software a utilizar para dar solución al problema planteado.

Capítulo 2: Características del sistema. Exploración, planificación y diseño: Se caracteriza la propuesta de solución. Se detallan las funcionalidades que tendrá la propuesta del sistema, lo que posibilitará su entendimiento. Se precisan las historias de usuario definidas, que recogen todas las funcionalidades a implementar. Se emplean los artefactos de la metodología de desarrollo en cada una de las etapas de desarrollo de la investigación.

Capítulo 3: Implementación y pruebas: Se expone todo lo relacionado a los procesos de implementación y pruebas del sistema.

CAPÍTULO 1: FUNDAMENTOS DE LA INVESTIGACIÓN, METODOLOGÍA Y TECNOLOGÍAS EMPLEADAS.

En el presente capítulo se realiza un análisis acerca de las herramientas que realizan vigilancia tecnológica a través de las técnicas de web scraping y recuperación de información, tanto a nivel internacional como nacional. Se describen las principales características de la metodología, las tecnologías, lenguajes y herramientas utilizadas para solucionar el problema de la investigación.

1.1 Marco conceptual

En el mundo actual, el gran avance de las tecnologías de comunicación con las herramientas de las Tecnologías de la información y la comunicación (TIC), hace que se cuente con demasiada información de dominio público donde puede consultarse, obteniéndose poco enfoque en lo que se quiere obtener o buscar, lo que conlleva a la inteligencia competitiva que es la práctica empresarial que reúne los conceptos y las técnicas que permiten articular el estudio del entorno, y a la cual aporta un gran valor la vigilancia tecnológica.

La VT sirve para anteponerse ante cualquier situación, se puede obtener un producto que pueda aplicarse a diferentes ámbitos abriendo nuevas oportunidades de negocio y diversificando a su vez dicho producto. Asimismo, ayuda a definir los planes estratégicos y, por lo tanto, identificar cuáles son las amenazas y las oportunidades de la organización. Además, gestiona el riesgo en función de la información que tenga la organización, contribuyendo a una mejor toma de decisiones.

Todo este proceso lleva un riesgo final, ya que una organización que quiere apostar por una tecnología, por un país, por un sector, debe saber que tiene un riesgo inherente a esa actividad y no es que se elimine el riesgo, sino lo que se hace es minimizarlo, controlarlo incluso hasta poder medirlo. (SOME ASPECTS REGARDING ON TECHNOLOGICAL VIGILA, 2014)

Una técnica de realizar vigilancia tecnológica es el *web scraping*, este es un término usado que describe el uso de un programa para extraer y recopilar datos de archivos HTML⁵ en la Internet. Normalmente, estos datos están en forma de datos con patrones, particularmente listas o tablas. (The Use of Web-scraping Software in Searching for Grey Literature, 2015)

Los *webs scrapers*⁶ son útiles en la búsqueda de información, de ejecución de acciones y toma de decisiones usando la información encontrada; o cuando un sitio web despliega información relevante, pero

⁵ HyperText Markup Language (Lenguaje de marcado por hipertexto)

⁶ Aplicaciones basadas en web scraping

no proporciona una interfaz de programación de aplicaciones (API) para que otras aplicaciones realicen consultas directas sobre dicha información. Estos sistemas ofrecen automatización y cierto nivel de inteligencia para la obtención y procesamiento de los datos. Podría decirse que permiten acceder a los datos camuflados en la capa de presentación, constituyendo una forma de adaptar el contenido web publicado a necesidades personales y capitalizar información que está aparentemente oculta. (Web Scraping, 2002 - 2014)

En cuanto al uso de esta información obtenida en este caso por los *webs scrapers* se utiliza la recuperación de la información evidenciándose mediante la representación de los elementos de la información, dando la posibilidad de almacenarlos para su organización y acceso. Esta idea no puede verse separada del usuario como elemento que valora la información que le brinda el sistema, dada la relevancia del resultado obtenido mediante su utilización. Para analizar este contenido recuperado desde la web existen varias técnicas que se pueden utilizar, como, por ejemplo:

- Técnica de ponderación⁷ de términos: Se basa fundamentalmente en el valor que poseen los términos dentro de la información, ya que algunos criterios poseen más valor que otros dentro de una misma búsqueda, por tanto, la ponderación pretende darle un valor adecuado a la búsqueda dependiendo de los intereses del usuario, los resultados obtenidos dependen en este caso del valor de dicha ponderación y este a su vez depende de los términos pertinentes que contenga el documento y la frecuencia con que se repitan. De forma que, el documento más pertinente de búsqueda sería aquel que tenga representado todos los términos de búsqueda y el valor más repetido, independientemente de donde se localice en el documento. (Mabrouk, y otros, 2017)
- Técnica de *clustering*: Es un modelo probabilístico que evalúa la frecuencia de los términos de búsqueda dentro de la información recuperada en forma de documentos. Se asignan pesos que agrupan los documentos por orden de importancia, mediante algoritmos de ranking. (Nicholls, y otros, 2018)
- Técnica de retroalimentación por relevancia: Obtiene el mayor número de documentos relevantes, después de aplicar varios criterios de búsqueda y de observar los documentos recuperados, se repite la consulta con elementos interesantes, seleccionados de los documentos primeramente recuperados. Utiliza un algoritmo genético para llevar a cabo la recuperación. (Demir, y otros, 2015)
- Técnica de *stemming*: Esta técnica elimina posibles confusiones semánticas que se puedan encontrar en la búsqueda de un concepto, debido a que las palabras morfológicamente están

⁷ Peso o relevancia que tiene algo

estructuradas por un prefijo y un sufijo, además de una raíz, se trunca la palabra y realiza la búsqueda solo por la raíz.

Se consigue el mayor aprovechamiento de esta técnica, mediante una correcta indexación⁸ en el proceso de tratamiento de los documentos, con ayuda de índices y tesauros (son un listado que presenta los términos que se utilizan para la representación de ciertos contenidos), permitiendo tener un listado terminológico controlado sobre un área o ámbito de conocimiento, que mantiene entre sí relaciones semánticas y genéricas, consiguiendo precisión terminológica en la búsqueda de información. De esta forma se evitan las ambigüedades léxicas y semánticas a la hora de establecer las consultas. (S.P.Ruba, y otros, 2015)

De las anteriores técnicas, para la realización del presente trabajo se podrían utilizar algunos de sus elementos en el tratamiento de la información recopilada; pero no se obtendría un resultado esperado, ya que faltaría un componente esencial, la interacción del usuario del sistema con los resultados que se obtengan.

La interacción de los usuarios con los sistemas de recuperación de la información (SRIs), pone de manifiesto la necesidad de los mismos de obtener información acertada. El usuario puede no estar satisfecho con la respuesta del sistema, evaluando el documento como no interesante o simplemente no accediendo a él.

El proceso donde el usuario expresa su satisfacción respecto al conjunto de bloques de información recuperada se denomina retroalimentación de relevancia. (Berzal, y otros, 2002) El sistema debe aprender de este proceso, por tanto, es preciso almacenar dicha información para mejorar la respuesta que da el sistema ante futuros accesos del usuario, dicha caracterización del usuario se almacena en el denominado perfil de usuario. Por esta razón es preciso utilizar una técnica que brinde la posibilidad de interrelación usuario-sistema, que concluya con un resultado de calidad que pueda ser utilizado por la Asesora de Mercadotecnia para contribuir a la vigilancia tecnológica y la toma de decisiones dentro del Centro TLM. Para lograr esta interrelación se emplea la Inteligencia Artificial (IA) a través de la lógica difusa.

La IA es una disciplina académica relacionada con la teoría de la computación cuyo objetivo es emular algunas de las facultades intelectuales en los sistemas artificiales. Se entiende por facultades intelectuales los procesos de percepción sensorial (visión, audición, entre otros) y a sus consiguientes procesos de reconocimiento de patrones, por lo que las aplicaciones más habituales de la IA son el tratamiento de datos y la identificación de sistemas. (Benítez, y otros, 2013) Una de las ramas de IA es la lógica difusa, la cual se utiliza para la resolución de una variedad de problemas.

⁸ Proceso mediante el cual se examinan ordenadamente los datos e informes para elaborar su índice

Lógica difusa

La lógica difusa es una técnica de la inteligencia computacional que proporciona una manera simple y elegante de obtener una conclusión a partir de información de entrada vaga, ambigua, imprecisa, con ruido o incompleta. En general la lógica difusa imita como una persona la toma decisiones basada en información con las características mencionadas. (Introducción al razonamiento aproximado: lógica difusa, 2006)

La lógica difusa en la recuperación de la información

La lógica difusa puede utilizarse como herramienta para mejorar los procesos de recuperación de información mediante pesos con representación difusa asociados a los términos (palabras), que forman documentos y consultas de usuarios, que pueden significar tanto presencia del término en el documento, como importancia del término en la consulta.

Se pueden definir cinco procesos diferentes relacionados con la recuperación de la información:

- El proceso de indexación, donde se genera la representación de los documentos.
- El proceso de consulta, donde el usuario expresa sus preferencias en lenguaje natural o mediante términos, que pueden estar conectados mediante operadores lógicos.
- El proceso de evaluación, donde se realiza una comparación de la representación de los documentos con la consulta del usuario, para determinar la relevancia de los mismos.
- El proceso de retroalimentación de relevancia, donde el usuario, de forma implícita o explícita expresa su satisfacción respecto al conjunto de documentos recuperados.
- El proceso de construcción y actualización del perfil de usuario, donde los términos que representan las preferencias del usuario, se almacenan con la finalidad de mejorar el comportamiento del SRI en futuros accesos del usuario al sistema.

Los SRIs tradicionales utilizan la comparación entre los documentos relevantes y los obtenidos luego de aplicar los modelos de espacio vectorial, probabilístico y booleano. Los términos de indexación de los documentos y los términos de las consultas se representan y evalúan de forma booleana mediante valores de presencia/ausencia, asociando valores o pesos numéricos a estos términos. (Berzal, y otros, 2002)

La lógica difusa es utilizada, dada la falta de flexibilidad, para las operaciones que manejan estos pesos en los modelos tradicionales, y la posibilidad de utilizar etiquetas lingüísticas para la interacción del SRI con el usuario. De esta forma al tener un conjunto de bloques de información $D = \{d_1, \dots, d_m\}$ y un conjunto de pesos asociados a los términos $T = \{t_1, \dots, t_n\}$, entonces SRI difuso se define mediante una función F como

se expresa a continuación (On the Delusiveness of Adapting a Common Space for Modeling IR Objects: Are Queries Documents?, 1993):

$$F : D \times T \rightarrow [0,1]$$

$$\forall d \in D, t \in T \quad F(d,t) = \mu_d(t)$$

Fig. 1 Función de SRI difuso.

En esta función, d es un bloque específico de información, D es un conjunto de bloques de información recuperado, t es un peso asociado a un término y T es el conjunto de términos presentes en los bloques de información recuperados. Se calcula una media de los pesos asociados a los términos presentes en cada bloque, del conjunto de todos los términos utilizados para obtener los bloques de información, este resultado queda clasificado utilizando etiquetas lingüísticas, las cuales quedan determinadas de la siguiente forma.

Grado de Pertenencia

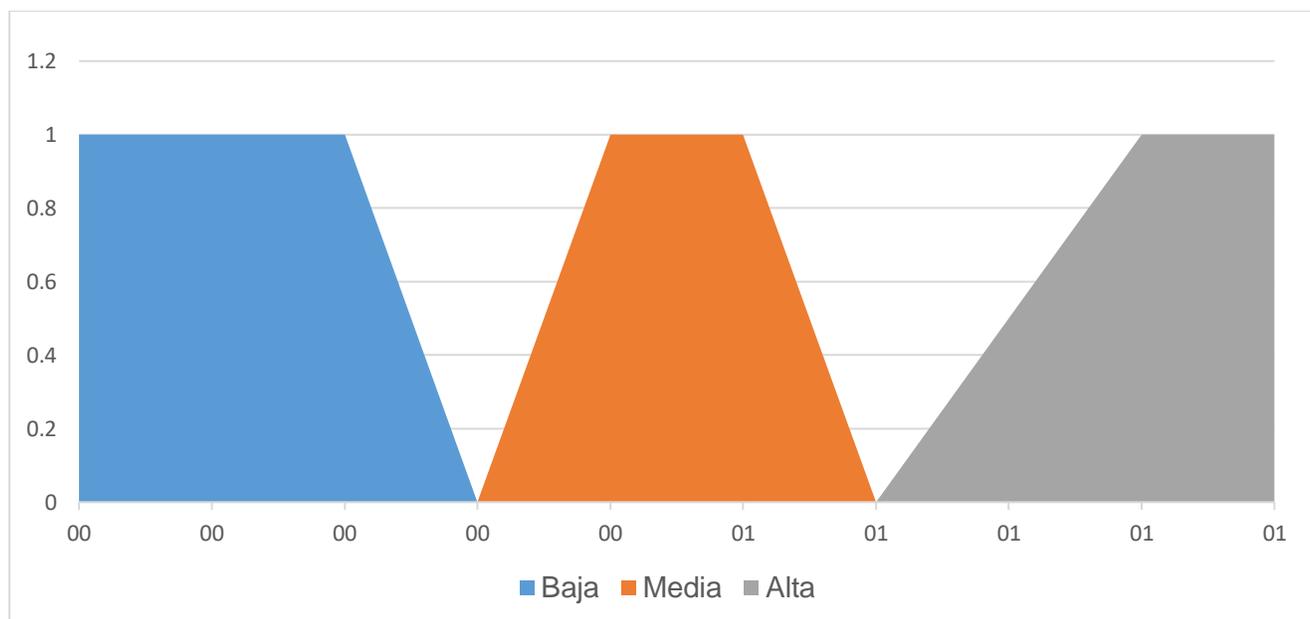


Fig. 2 Resultado de la función para cada bloque de información.

Para definir los umbrales de clasificación para los bloques de información, se utilizó la función para estudiar junto al cliente los resultados que se arrojaban al clasificar información que le era de importancia. Se identificaron 30 bloques de información de los cuales 12 resultaron tener un valor entre $[0.1; 0.3]$ como resultado de la función del SRI difuso, a los cuales el cliente identificó de baja importancia para sus necesidades de información. Se identificaron 9 bloques de información con valores entre $[0.3; 0.6]$

clasificados por el cliente como de media importancia y 9 bloques de información con resultados entre [0.6; 1] a los que el cliente identificó como de alta importancia para sus necesidades de información. Junto a este estudio realizado al cliente, donde su valoración es considerada como un juicio de experto (Robles Garrote, y otros, 2015), se consultó además una extensión del modelo booleano mediante etiquetas lingüísticas para determinar la relevancia de los bloques de información.

Proceso de retroalimentación a través del usuario.

La acción que ejerce el usuario de evaluar según sus necesidades los resultados que arroja el sistema puede definirse como retroalimentación, dada la respuesta del sistema a las diferentes consultas, se pueden tener otros parámetros donde se evalúan las preferencias del usuario sobre una lista de documentos.

Retroalimentación explícita:

Se basa en la información que recibe el usuario sobre la respuesta que le brinda el sistema luego de realizar una determinada consulta. Se puede utilizar esta evaluación del usuario para que el sistema mejore la clasificación de los documentos, además construya y mantenga los perfiles de usuario con sus necesidades informativas específicas.

- Selección de términos para mejorar consultas.

Se realiza este tipo de retroalimentación cuando los términos de la consulta no arrojaron los resultados esperados por el usuario, haciendo una reevaluación de la consulta con nuevos términos obtenidos de los documentos con mayor grado de aceptación del conjunto inicial.

El usuario puede asignar un grado de importancia tanto a los términos que aparecen originalmente en la consulta, como los provenientes de la lista de candidatos de términos a añadir a la consulta. Dicho grado también es conveniente que se exprese lingüísticamente, ya que para el usuario es difícil establecer un grado numérico. (Bordogna, y otros, 1995)

- Evaluación de documentos del conjunto recuperado.

Teniendo unas listas de documentos que el sistema previamente haya evaluado, se procede a la interacción con el usuario al mostrarle los mismos junto a su evaluación, y así, el usuario puede realizar una propia decisión, determinando el de más importancia, según su criterio. (Yao, 1995)

La lógica difusa en el proceso de construcción de perfiles de usuario.

Los perfiles de usuarios se pueden definir como la asociación de un conjunto de términos y un peso ligado a ellos, que indica el interés del usuario en los temas relacionados con ese término. Los mismos pueden

ser extraídos a partir de consultas previas y de términos de indexación de los documentos relevantes para dichas consultas. (Bordogna, y otros, 1995)

Dados los diferentes productos del Centro TLM de la Universidad de Ciencias Informáticas, sobre los que se realizará la búsqueda de información, como parte del proceso de vigilancia tecnológica, se podrá identificar a los mismos como perfiles, donde se le asocien términos definidos por selectores, que brindarán la posibilidad de filtrar la información y clasificarla por nivel de importancia para el usuario, utilizando la lógica difusa en el proceso de recuperación y tratamiento de la información obtenida mediante el *web scraping*.

Por lo descrito anteriormente, tanto la técnica de *web scraping* como la recuperación de la información mediante el uso de la lógica difusa, deben constituir una parte esencial de la solución del problema de la investigación en curso. Sin embargo, por sí solas no constituyen una solución, presentando algunas particularidades que impiden su uso directo:

- Se debe programar manualmente el acceso a cada una de las páginas de donde se desea obtener la información; así como, la lógica de procesamiento y/o reutilización de la misma.
- El proceso de parsear⁹ el HTML para obtener la información de interés, puede ser una tarea difícil teniendo en cuenta que la estructura de la página va a depender de los lenguajes, las tecnologías y herramientas seleccionadas por los desarrolladores.
- Se debe tratar de no afectar el desempeño del sitio que se analiza, evitando ataques de denegación de servicio no deseados.
- Es necesaria la interacción con el usuario para obtener mejores resultados a medida que el sistema vaya aprendiendo que información le es más relevante para las diferentes consultas que se le realicen.

Por estas razones, se analizaron aplicaciones que obtienen información desde la web, para recopilar información publicada, que permitieran emplear las potencialidades del *web scraping* y la recuperación de la información mediante la técnica de lógica difusa sin necesidad de programar manualmente todo el proceso.

⁹ analizar el código

1.2 Consumo y reutilización de contenido web. Estado del arte

Hoy día existen diversas aplicaciones para la extracción de contenido web, las cuales varían en dependencia a sus funcionalidades y los precios. Atendiendo a lo anterior, se lleva a cabo un estudio sobre estas:

Herramientas a nivel internacional

1.2.1 Import.io

Es una plataforma gratis basada en la nube, esto significa que no necesita ejecutar *web scraping* en la computadora y todos sus datos se mantienen en la nube. Creada en el 2012, tiene como objetivo convertir la web en una base de datos. Permite integrar su base de datos con otro software ofreciendo una API, que brinda la búsqueda y recuperación de datos desde un almacén. El software está diseñado para los usuarios no técnicos, intentando "adivinar" lo que necesitan. Tiene la posibilidad de ajustarse manualmente a sus necesidades. (Web Scraping, 2002 - 2014)

La aplicación anterior a pesar de cumplir con varias funcionalidades para la extracción de contenido web y ser gratis, no puede ser utilizada por el inconveniente siguiente:

- No se puede controlar el proceso de raspado avanzado. No hay funciones de control adicionales para ajustar filtros.

1.2.2 Helium Scraper

Es una herramienta de extracción de datos permanente. Este extractor de datos utiliza un algoritmo de búsqueda para raspado que asocia los elementos a extraer por sus propiedades HTML. Facilita la extracción y manipulación de información más compleja con la ayuda de secuencias de comandos de JavaScript y SQL. Es bueno para las estructuras de datos visuales de múltiples niveles. (Web Scraping, 2002 - 2014)

A pesar de cumplir con todas las cualidades anteriormente mencionadas, no es rentable ya que es un software privativo y presenta las siguientes desventajas:

- Pago de licencia: Helium Scraper presenta 4 licencias, la de Basic tiene una duración de 3 meses y con un costo de \$ 99.00 USD. La Professional tiene una duración de 6 meses y un costo de \$199.00 USD. La licencia Business tiene una duración de 12 meses y un costo de \$ 399.00 USD; y la Enterprise es por 2 años y un monto de \$ 699.00 USD.
- Solo es compatible con Windows.

1.2.3 Web Data Extractor

Funciona haciéndole scrap a una lista de URLs¹⁰ de un archivo, o mediante una palabra clave en los motores de búsqueda. Puede permitir que siga vínculos externos desde las páginas originales, con la capacidad de ir tan profundamente en las rutas URL que necesite. Mediante el uso de funciones recursivas, es posible realizar búsquedas en todo el Internet parándolas manualmente. (WebExtractor System, 2018)

Aunque es una herramienta para obtener información de las páginas web, su utilización no es posible debido a que:

- Las páginas web que detectan este scan, lo toman como una forma invasiva por su recursividad, por lo que pueden llegar a bloquear las direcciones desde donde se realiza la búsqueda.
- Es solo compatible con Windows.
- Es un software privativo con un costo de \$89.00 a \$199.00 USD.

1.2.4 OutWit Hub Review

Es un software que proporciona una simple extracción de datos. Permite obtener de las páginas web enlaces, texto, e imágenes. Permite al usuario elegir lo que necesita. El programa es apto para navegar por los enlaces en las páginas, por lo que esta característica funciona bien si se requiere varias cadenas de *web scraping*. Presenta una funcionalidad para evitar el bloqueo de servidores, haciéndole un retraso configurable a las peticiones que hace. Además, presenta una versión libre de costo. (Web Scraping, 2002 - 2014)

Es un software con buenas posibilidades para el *web scraping*, pero presenta dificultades por lo que no es conveniente utilizarla por las siguientes razones:

- Presenta versiones diferentes, con carácter comercial presenta la Pro con un costo de \$80.91 USD.

1.2.5 WebHarvy

Es un extractor de contenido web en el que se puede hacer configuraciones para lo que se necesite buscar en las páginas, presentando la posibilidad de aplicar filtros avanzados. Presenta funcionalidades para sitios web de respuesta lenta o sitios que emplean AJAX¹¹ y además presenta un Proxy configurable. Se puede

¹⁰ Uniform Resource Locator (localizador de recursos uniforme)

¹¹ Asynchronous JavaScript And XML (JavaScript asíncrono y XML)

exportar los datos capturados como un archivo (archivo XML¹², CSV¹³ o TSV¹⁴) o insertarlo en una base de datos con la acción de 'Exportar'. Este extractor es estable y fácil de usar y fácil de aprender. (Web Scraping, 2002 - 2014)

Aun realizando el raspado rápido de la web en varios formatos de datos de exportación, no se puede utilizar por las siguientes razones:

- Solo es compatible con el sistema operativo Windows.
- Es un software privativo; posee una licencia de prueba de duración de 30 días y un costo de \$99.00 USD.
- La búsqueda de la información se hace sobre páginas web abiertas en el momento a través de la aplicación.
- No presenta la posibilidad de hacer búsqueda programadas o de definir direcciones web para la búsqueda en segundo plano.

1.2.6 Octoparse

Es un software de extracción de datos web. Proporciona a los usuarios una interfaz de usuario punto a punto para desarrollar patrones de extracción, de modo que los raspadores pueden aplicar estos patrones a sitios web estructurados. Brinda numerosas funcionalidades a los usuarios como por ejemplo: el bloqueo de anuncios para extraer datos de páginas, donde se optimizará el tiempo de carga de la página y se reduce el número de solicitudes HTTP¹⁵. Puede hacer varias tareas configurables para la extracción de los datos, además brinda la posibilidad de hacerlo local en la propia PC. (Octopus Data Inc, 2017)

Aunque es una aplicación de las más completas en cuanto a procesamiento de datos, no se puede utilizar por las siguientes limitantes:

- Presenta versiones diferentes de carácter comercial Free, Estándar y Profesional, donde la versión libre presenta funcionalidades limitadas.
- Exporta los datos en formato CSV, Excel, HTML, TXT, y hacia bases de datos en MySQL, SQL Server, Oracle.

¹² Extensible Markup Language (Lenguaje de marcado extensible)

¹³ comma-separated values (valores separados por coma)

¹⁴ Tab Separated Values (Valores separados por tabulaciones)

¹⁵ Protocolo de transferencia de hipertexto

1.2.7 SoftVT

Es una herramienta creada para automatizar aquellos procesos relacionados con la captación, gestión y difusión de información estratégica (tecnológica, mercado, económico-comercial, competencia, entre otros.). Permite procesos tanto de vigilancia tecnológica y económica como otros más centrados en el ámbito de la inteligencia competitiva. Es adaptable a diferentes necesidades y realiza una selección de servicios en función de los requisitos que defina el usuario. (AIMPLAS, 2017)

1.2.8 Vigiale

Es una Plataforma Web para la gestión de la VT y del entorno que permite el seguimiento de fuentes de información seleccionadas y la notificación de cambios detectados. Permite gestionar de forma integrada distintas fuentes de información, ordenarlas, clasificarlas y actualizarlas, utilizando tecnologías modernas de captura, categorización, indexación y filtros de diversa naturaleza, personalizable en función de las necesidades y requerimientos de cada empresa u organización. (Tecnología IALE, 2016)

1.2.9 Hontza

Es una herramienta para la inteligencia competitiva y estratégica desarrollada en código abierto, con licencia GPL¹⁶. Hontza permite la gestión integral del proceso de la vigilancia tecnológica y la inteligencia competitiva, con un enfoque orientado más al grupo que al usuario individual. Se basa en un estilo colaborativo de trabajo, y en el análisis humano de la información. Además, su licencia GPL permite que, si la empresa dispone de personal capacitado, pueda adaptar y modificar la herramienta libremente según sus necesidades. (CDE)

Las últimas tres aplicaciones utilizan tecnologías y herramientas que brindan un resultado ajustado a las necesidades que se describen en este documento para darle solución al problema de investigación, pero presentan inconvenientes, el principal de ellos es que son extranjeras, cobrando así los servicios brindados, en el caso de Hontza cuenta con un período de prueba gratis online de 15 días, luego del cual se tendría que pagar licencia para su uso, y para el trabajo con las herramientas SoftVT y Vigiale es necesario establecer un contrato con sus desarrolladores para utilizar los servicios online que ofrecen.

Herramienta a nivel nacional

¹⁶ General Public License (licencia pública general)

1.2.10 Web Scrap

Se encontró la investigación de Luis Manuel Goncalves Torres, titulada: “Herramienta de consumo y reutilización de contenido web”. Esta investigación hace referencia a un sistema que permite informatizar el proceso de consumo y reutilización de contenido web en la UCI. Esta aplicación está limitada ya que solo realiza consumo de web sin utilizar recuperación información usando la lógica difusa. (Gonçalves Torres, 2015)

Conclusiones del estado del arte

Con el análisis de todas las aplicaciones anteriores, se concluye que las mismas no resuelven la problemática planteada en la presente investigación, atendiendo a que el 60% de las aplicaciones son privativas, costosas o creadas para un solo sistema operativo. Se hace necesario desarrollar una aplicación informática que realice *web scraping* y recuperación de información usando la lógica difusa, respondiendo a las necesidades del proceso de Mercadotecnia del Centro TLM, que incluya las funcionalidades comunes de los sistemas analizados anteriormente.

1.3 Metodología

El equipo de desarrollo está compuesto por tres personas (dos desarrolladores y el cliente), se acordó la entrega al cliente de una aplicación funcional en poco tiempo, lo que requiere que la documentación que se genere sea la imprescindible.

Atendiendo a lo anterior, se analizaron los enfoques acerca de las metodologías y prácticas más convenientes en el proceso de desarrollo de software, utilizando el método Boehm y Turner que permite caracterizar el proyecto de software a partir de 5 criterios y estimar cuán ágil o prescriptivo debería ser el enfoque a utilizar. (Elección entre una metodología ágil y tradicional basado en técnicas de soft computing, 2016) Los criterios a evaluar son:

Tabla 1. Criterios para la selección apropiada de la metodología

Factor	Discriminadores ágiles	Discriminadores formales
Tamaño	<ul style="list-style-type: none"> ➤ Escalabilidad y conocimiento. ➤ Dependencia y escalabilidad limitada por el porcentaje alto de conocimiento tácito. ➤ Apropriado para equipos y productos pequeños. 	<ul style="list-style-type: none"> ➤ Escalabilidad y conocimiento explícito. ➤ Apropriado para productos y equipos grandes. ➤ Difícil de mantener en pequeños proyectos.

Criticidad	La simplicidad en la documentación y el diseño dificulta los planes de pruebas. No aconsejado para sistemas con niveles de criticidad altos (IEEE 1012).	<ul style="list-style-type: none"> ➤ Rigor de requisitos y diseño adecuados para procesos de pruebas, verificación y validación. ➤ Difícil de gestionar en proyectos de escasa criticidad
Dinamismo	“Refactorizar” desde un diseño básico hasta el producto final es un método ideal para entornos dinámicos e innovadores, pero muy caro por el retrabajo para entornos estables o conocidos.	En sistemas estables y conocidos, a partir de requisitos completos y diseños detallados permite trazar y seguir un plan completo y “hacerlo bien a la primera”.
Personal	Los métodos de trabajo ágiles requieren una masa crítica de técnicos con niveles de experiencia medios o altos, capaces de comprender y adaptar los métodos y las técnicas empleadas. Consiste en grupos pequeños y trabajando en el mismo local.	Aunque es aconsejable contar con personas expertas en las fases de definición del proyecto, luego pueden ejecutarse con menor masa crítica de expertos. Consiste en grupos grandes y pueden trabajar de forma distribuidas.
Cultura	Más apropiado para culturas de <i>empowerment</i> responsabilidad y horquilla de decisión y libertad personal.	Más apropiado en culturas en las que las personas se sienten seguras con un marco de tareas y responsabilidades bien definido.

La siguiente Fig. 3 muestra los factores aplicados a las condiciones del proyecto. Nótese que todos los parámetros sugieren el uso de una metodología ágil:

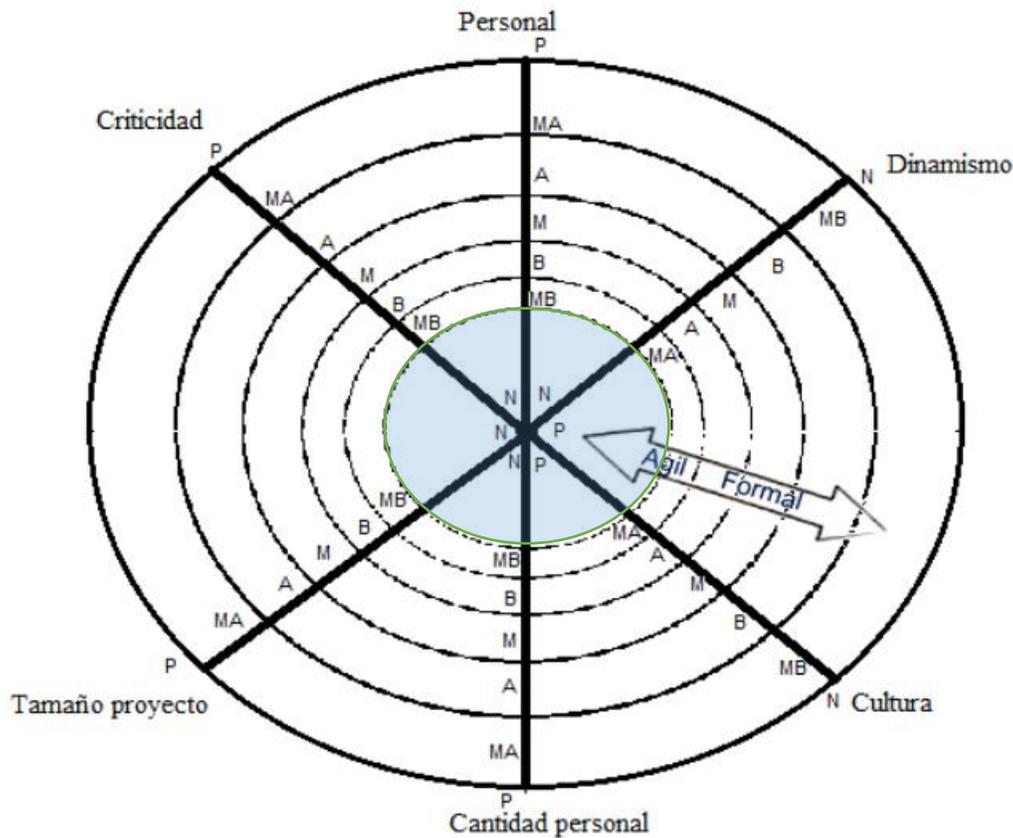


Fig. 3 Resultados del empleo de los criterios.

Se escoge la metodología ágil *Extreme Programming* (XP) teniendo en cuenta que:

- Se aplica de manera dinámica durante el ciclo de vida del software.
- Es capaz de adaptarse a los cambios de requisitos.
- Los individuos e interacciones son más importantes que los procesos y herramientas: La gente es el principal factor de éxito de un proyecto software. Es más importante construir un buen equipo que construir el entorno.
- Es más importante desarrollar un software que funcione que documentación exhaustiva: La regla a seguir es no producir documentos a menos que sean necesarios de forma inmediata para tomar una decisión importante. Estos documentos deben ser cortos y centrarse en lo fundamental.

- La colaboración con el cliente es más importante que la negociación de contratos: Se propone que exista una interacción constante entre el cliente y el equipo de desarrollo. Esta colaboración entre ambos será la que marque la marcha del proyecto y asegure su éxito.
- La respuesta ante el cambio es más importante que el seguimiento de un plan. (Bustamante, y otros, Marzo del 2014)
- Algunas prácticas de XP como las entregas pequeñas, el diseño simple y el Desarrollo Guiado por Pruebas (Test Driver Development) son parte de la filosofía de desarrollo del equipo. (Cockburn, 2000)

1.4 Tecnologías y herramientas de desarrollo

Para el desarrollo de la solución se utilizaron varias tecnologías y herramientas teniendo en cuenta cada una de sus características esenciales. A continuación, se describen brevemente las mismas:

1.4.1 Python 2.7.14

Python es un lenguaje de programación de propósito general con un entorno con alta simplicidad, versatilidad y rapidez de desarrollo. Es un lenguaje de *scripting* independiente de plataforma y orientado a objetos donde se pueden desarrollar aplicaciones para Windows, servidores de red y páginas web. Al ser un lenguaje interpretado no es necesario compilar el código fuente para ejecutar sus aplicaciones, dando la ventaja de desarrollo rápido e inconvenientes de menor velocidad de procesamiento y ejecución. Su uso ha aumentado en los últimos años dados que presenta un

La cantidad de librerías con numerosas tareas habituales, las que pueden ser utilizadas sin ser programadas desde cero. La sencillez de su código, el cual ahorra líneas de código frente a otros lenguajes como Java o C. (Python Software Foundation (US) , 2001 - 2018)

Este lenguaje posee ventajas para el desarrollador ya que facilita el trabajo con librerías al no tener que implementar funciones que ya están integradas. Algunas librerías de este lenguaje, como Request y Beautiful Soup, permiten interactuar con las páginas web a consultar y también son utilizadas para la conexión y recopilación de datos desde los sitios, siendo esto una función necesaria en la aplicación.

1.4.2 Django 1.11.6

Django es un *framework*¹⁷ de alto nivel escrito en Python, *open source*¹⁸ y gratuito que brinda la posibilidad de desarrollo de aplicaciones web de una forma rápida, limpia y pragmática. Hereda las funcionalidades de Python, entre ellas escribir código bastante fácil de entender, y sobre todo permite desarrollar aplicaciones muy rápidas y potentes. Presenta la posibilidad de utilizar ORM¹⁹ la cual permite realizar consultas SQL a la Base de Datos sin usar SQL, ya que crea una base de datos orientada a objetos virtual, sobre la base de datos relacional, con la cual se puede trabajar con el modelo orientado a objetos utilizado por Python. (Django Software Foundation, 2005-2018)

Se empleó este *framework* ya que permite al programador realizar las tareas en un plazo de tiempo relativamente corto, así como utilizar la numerosa información disponible. Dado el trabajo que es necesario realizar con los datos se hace imprescindible las funcionalidades que brinda el ORM, pues los datos deben ser consultados de una manera rápida y sencilla para no utilizar demasiados recursos en esa funcionalidad.

1.4.3 JetBrains-PyCharm 2017 2.2

PyCharm es un IDE²⁰ propietario de lenguaje específico para desarrolladores Python y Django, desarrollado por JetBrains que brinda una licencia gratuita para fines educacionales. Contiene un editor de código inteligente, el cual interpreta los detalles específicos de Python, así como mejora la efectividad al hacer uso de formateo automático y completamiento de código, refactorizaciones e importación automática además de otras funcionalidades. Estas características respaldadas por avanzadas rutinas de análisis de código hacen de PyCharm una potente herramienta de desarrollo en Python. (JetBrains s.r.o. , 2000 - 2018)

Esta herramienta es de fácil utilización para los desarrolladores que decidan trabajar en aplicaciones web como de escritorio por la dinámica de su interfaz. Sus funcionalidades permiten trabajar directamente con las rutinas de ejecución del servidor web de Django, así como la utilización de sus diferentes comandos desde la propia aplicación, lo que lo hace aún más potente para este trabajo.

1.4.4 Beautiful Soup 4.6

Esta es una biblioteca escrita en Python diseñada para implementar aplicaciones que realicen *web scraping* algunas de sus características más interesantes se exponen a continuación (Richardson, 1998-2018):

¹⁷ Español: marco de trabajo

¹⁸ Español: fuente abierta

¹⁹ Mapeo objeto-relacional

²⁰ Entorno de Desarrollo Integrado

- Provee métodos robustos para navegación, búsqueda y modificación del árbol DOM²¹ de cualquier documento HTML, así como una completa caja de herramientas para extraer exactamente la información que se desea.
- BeautifulSoup realiza automáticamente las tareas de codificación y decodificación necesarias en el proceso de *web scraping*.
- Se integra fácilmente con el analizador sintáctico²² de HTML para Python más utilizados, como lxml y HTML5lib, permitiendo probar diferentes estrategias en cuanto a velocidad y flexibilidad. (Richardson, 1998-2018)

Las características que posee permiten que se manipulen a conveniencia del usuario la estructura del documento de la que se quiere obtener la información haciendo que el trabajo sea mucho más sencillo para los desarrolladores, por lo que es una librería de vital importancia para esta investigación.

1.4.5 Requests 2.7.0

Requests es una librería para HTTP escrita en Python y licenciada bajo Apache2. Esta elimina las complicaciones de trabajar HTTP en Python haciendo que la integración con servicios web sea transparente quitando la necesidad de agregar *queries*²³ manualmente a las direcciones URL, o convertir la información a formularios para hacer una petición POST a las misma. (A Kenneth Reitz Project., 2013) Esta librería presenta funcionalidades que permiten a los desarrolladores un uso sencillo de las conexiones HTTP necesarias para que se realice la recopilación de la información desde los sitios web.

1.4.6 HTML5

HTML5 puede ser utilizado con tan solo un navegador web moderno. Las tecnologías móviles también presentan navegadores que son perfectamente compatibles con HTML5. La nueva generación de HTML reemplaza a HTML4.01 x HTML 1.0 y x HTML 1.1 proveyendo de nuevas características necesarias para las aplicaciones web modernas como presentar una lona de dibujo de dos dimensiones que puede ser programada con JavaScript, soporte para video que puede ser empotrado dentro de las páginas web sin la necesidad de *plugins* de terceros, almacenamiento local persistente sin necesidad de *plugins* externos, aplicaciones web que pueden seguir funcionando luego de quedar offline y soporte para geolocalización que los usuarios pueden utilizar para conocer su posición utilizando la propia aplicación. (Pilgrim, Agosto 2010)

²¹ *Document Object Model* (modelo de objetos de documentos)

²² Buscar

²³ Español: consultas, preguntas

1.4.7 CCS3

Hojas de Estilo en Cascada (Cascading Style Sheets) es el lenguaje utilizado para describir la presentación de documentos HTML o XML, esto incluye varios lenguajes basados en XML como son X HTML²⁴ o SVG²⁵. CSS describe cómo debe ser renderizado el elemento estructurado en pantalla. Es uno de los lenguajes base de la *Open Web* y posee una especificación estandarizada por parte del W3C²⁶. Desarrollado en niveles, CSS1 es ahora obsoleto, CSS2.1 es una recomendación y CSS3, ahora dividido en módulos más pequeños, está progresando en camino al estándar. (Mozilla, y otros, 2005-2018)

1.4.8 JavaScript (JS)

Es un lenguaje ligero e interpretado, orientado a objetos con funciones de primera clase, más conocido como el lenguaje de script para páginas web, pero también usado en muchos entornos sin navegador, tales como node.js o Apache CouchDB. Es un lenguaje script multi-paradigma, basado en prototipos, dinámico, soporta estilos de programación funcional, orientada a objetos e imperativa. (Mozilla, y otros, 2005-2018)

1.4.9 jQuery 1.9.11

Es una biblioteca de JavaScript rápida, pequeña y rica en funciones. Realiza recorridos y manipulación de documentos HTML, manejo de eventos y animación. Presenta Ajax, para realizar peticiones HTML dinámicamente la cual es fácil de usar y funciona en una multitud de navegadores. Con una combinación de versatilidad y extensibilidad. (The jQuery Foundation, 2018)

1.4.10 Bootstrap 3

Es un kit de herramientas de código abierto para desarrollar con HTML, CSS y JS. Haciendo un prototipo rápido de las ideas y permitiendo la creación de aplicaciones completa, sistema de cuadrícula sensible, componentes precompilados extensos y potentes complementos basados en jQuery. Es una biblioteca que facilita el trabajo a los desarrolladores ya que tiene incluidos diversos estilos y funcionalidades que son necesarias para realizar un sitio web de calidad, dado lo estándares actuales. (Otto)

1.4.11 RSS²⁷

Los RSS son un formato de archivos basados en XML, estandarizado para distribuir contenidos, usualmente las novedades de los sitios web, por un canal distinto de la propia página web y con una dirección específica que es apuntada desde la página web principal. Por la simplicidad que posee la estructura del archivo RSS, se hace fácil la recopilación de la información contenida utilizando *web scraping*, obteniéndose las nuevas

²⁴ *eXtensible HyperText Markup Language (Lenguaje de marcado de hipertexto extensible)*

²⁵ *Scalable Vector Graphics (Gráficas vectoriales escalables)*

²⁶ *World Wide Web Consortium (El Consorcio World Wide Web)*

²⁷ *Really Simple Syndication (Sindicación realmente simple)*

publicaciones que se hacen en los sitios, siendo una fuente actualizada que puede ser utilizada para recuperar la información. Por lo que la recuperación de las URLs que brindan los canales RSS y la recuperación de su contenido son funcionalidades importantes para la aplicación. (Wusteman, 2004)

Las tecnologías anteriormente mencionadas son importantes durante el desarrollo de la aplicación, ya que algunas son de las más avanzadas en el mundo de la informática actual y se ajustan perfectamente a las necesidades que requiere la implementación de las funcionalidades de la aplicación a desarrollar.

1.4.12 PostgreSQL 9.4

PostgreSQL es un sistema gestor de bases de dato objeto-relacional, distribuido bajo licencia BSD y con su código fuente disponible libremente. Este sistema es el más potente del mercado en comparación a otros de código abierto además de ser una fuerte competencia para otros gestores de base de datos comerciales. PostgreSQL funciona muy bien con grandes cantidades de datos y una alta concurrencia de usuarios accediendo a la vez al sistema. Además de presentar una excelente integración con Python como lenguaje a utilizar, así como la mejor compatibilidad con el ORM de Django. (Martinez Guerrero, 2009-2013)

La utilización de este gestor de base de datos en el desarrollo del sistema informático está fundamentada en lo planteado anteriormente, brindando una eficiente y segura forma de gestionar la información generada desde la aplicación.

1.4.13 Visual Paradigm 15.0 Community Edition

Visual Paradigm es una herramienta CASE: Ingeniería de Software Asistida por Computación. La misma propicia un conjunto de ayudas para el desarrollo de programas informáticos, desde la planificación, pasando por el análisis y el diseño, hasta la generación del código fuente de los programas y la documentación. Las funcionalidades de generar estos diagramas le permiten hacer la ingeniería inversa de bases de datos desde Sistemas Gestores de Bases de Datos (DBMS) existentes y crear diagramas de Entidad-Relación. Por estas características se utilizó para modelar el diagrama de Entidad-Relación resultante de la base de datos creada por el ORM de django para esta aplicación. (Pressman, 2005)

1.5 Conclusiones del capítulo

A través de la investigación para la conformación del marco teórico se consiguió organizar, estructurar y guiar el trabajo. Además, se identificaron y analizaron aplicaciones que realizan vigilancia tecnológica, concluyéndose en desarrollar un sistema que realice vigilancia tecnológica mediante las técnicas de web scraping y recuperación de información con lógica difusa que resuelva el problema planteado. Asimismo,



se seleccionaron las tecnologías, herramientas, los lenguajes y la metodología de desarrollo a emplear en la solución propuesta a desarrollar.

CAPÍTULO 2. CARACTERÍSTICAS DEL SISTEMA. EXPLORACIÓN, PLANIFICACIÓN Y DISEÑO.

En este capítulo se abordarán los temas relacionados con la fase de exploración, planificación y diseño de la metodología de desarrollo XP, donde se describen las características del sistema a desarrollar y su arquitectura. Se describen las Historias de Usuarios (HU) que proporcionan un mayor entendimiento y comprensión del sistema y se elabora la estimación del esfuerzo que costará implementar cada una de estas. Se identifican los patrones de diseños y se confeccionan las tarjetas CRC. Además, se describe el algoritmo a desarrollar.

2.1 Características de la propuesta del sistema

Se decidió implementar una aplicación que, a partir de un control de acceso, definido por roles donde los usuarios posean determinados permisos administrativos y puedan realizar un monitoreo de información relevante para ellos desde diferentes sitios de internet. La aplicación permite agregar temas de búsqueda, estos quedarán registrados en el sistema como perfiles de usuario. Estos perfiles, presentarán las diferentes recopilaciones de documentos que realice el sistema luego de una búsqueda, quedando clasificados en términos lingüísticos que permitan una diferenciación (buena, regular, mala) por métodos definidos dentro del sistema y brindando la posibilidad de interactuar con el usuario al permitirle hacer su propia evaluación de lo que ha sido indexado.

La recopilación de la información se hará a partir de diferentes direcciones URLs que el usuario puede agregarle al sistema de dos maneras posibles, tanto incluyéndola directamente el mismo o seleccionándola de una lista de posibles sitios relacionados con alguna(s) palabra(s) clave(s) que estime conveniente, también presenta la funcionalidad de eliminar estas direcciones URLs o modificarlas de acuerdo a sus necesidades.

En cuanto a la realización de la selección de la información realmente relevante dentro de los diferentes sitios, se realizará a partir de selectores, que no serían más que términos claves para el usuario dentro de la información que desea de cada sitio. La calidad de los textos obtenidos se obtendrá por una función booleana de presencia de los selectores, dando una solución fuzificada²⁸, definidas por términos lingüísticos. El usuario puede hacer una inclusión de los selectores, así como una eliminación y modificación de los ya existentes. El inicio del monitoreo de la información podrá hacerlo el usuario una vez tenga definidas las direcciones y los selectores por temática, este monitoreo podrá ser de manera manual o automática, permitiendo con esta última programar un monitoreo en una fecha y hora específica.

²⁸ Conversión de un valor numérico de medida en una variable lingüística

Luego de que se obtenga la información deseada, el usuario puede generar un informe con la misma, posibilitando el sistema de enviarla como adjunto en un correo electrónico o en el propio cuerpo del correo, este envío podrá ser configurado de manera automática para que sin la presencia del usuario el sistema pueda enviar informes en un momento especificado por el usuario.

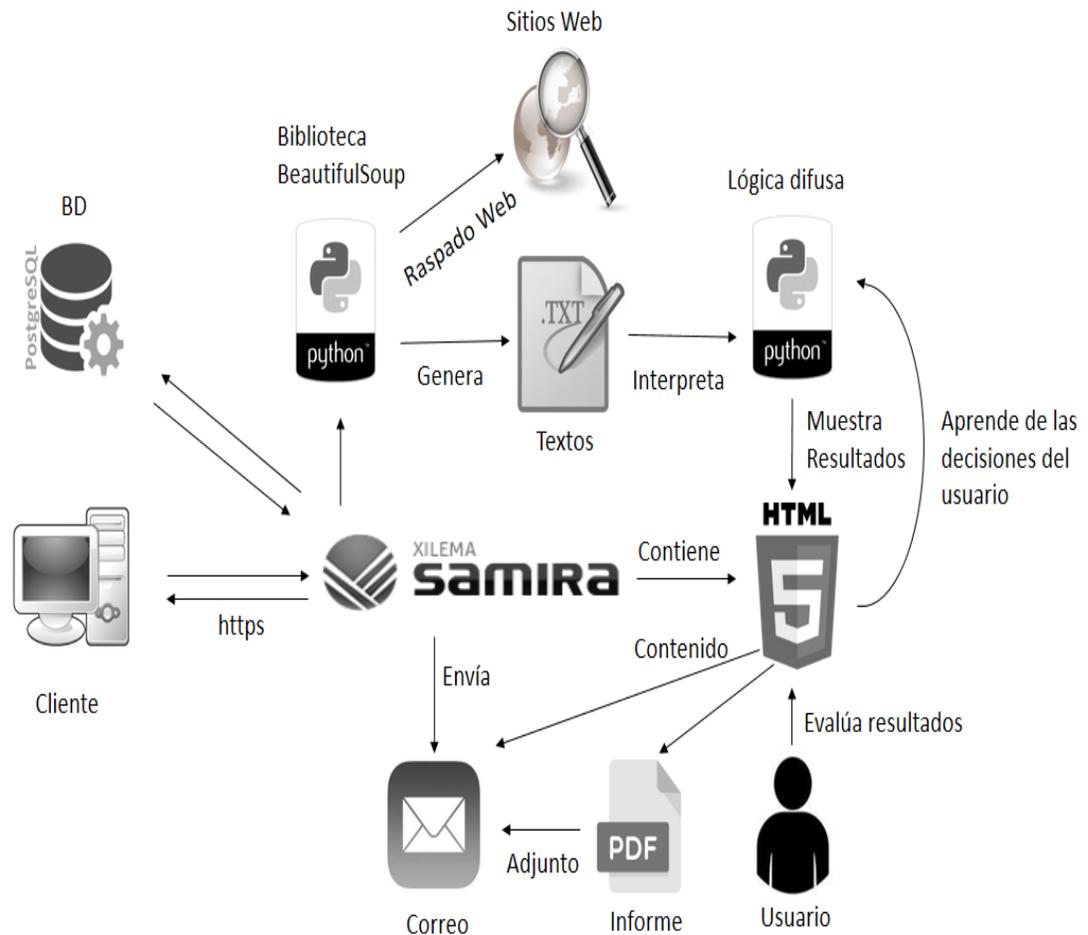


Fig. 4 Propuesta del sistema.

2.2 Lista de reservas de producto

Las características funcionales a desarrollar en el sistema son:

- Autenticar usuario.
- Gestionar las temáticas asociadas a los productos del centro.
- Gestionar las palabras claves a emplear en la búsqueda de información.
- Gestionar las URLs sobre las cuales buscar.

- Iniciar y detener monitoreo de URL.
- Configurar automáticamente el monitoreo de las URLs seleccionadas.
- Listar el registro de la búsqueda y el número de URLs analizadas.
- Gestionar los resultados de las búsquedas obtenidas de las URLs analizadas.
- Generar documento en formato PDF con resultados seleccionados.
- Configurar el envío automático de los resultados obtenidos de las URLs analizadas.
- Configurar el envío por correo de los resultados obtenidos.

2.3 Características no funcionales del sistema

Las características no funcionales del sistema son propiedades o cualidades que el producto debe tener, para así, hacer al producto atractivo, usable, rápido o confiable. (Cockburn, 2000) Las presentes en la aplicación son:

Seguridad

- **Integridad:** Mediante el rol de superusuario se garantiza que este sea el único que puede modificar la información obtenida de las URLs analizadas, mientras que los demás usuarios solo tendrán permiso de lectura de la información guardada en la base de datos, que se protegerá a través de la integridad referencial que brinda el gestor de bases de datos PostgreSQL, manejada por el sistema asegurando la coherencia de los datos existentes.
- **Disponibilidad:** Los datos estén disponibles para su consulta en cualquier momento. La aplicación guardará los datos dentro de su propia estructura, pudiendo así consultarlos en cualquier momento deseado. El gestor de base de datos deberá contar con un sistema de respaldo capaz de realizar una salva de la base de datos cada cierto tiempo para en caso de error o daños poder restaurarla.
- **Confidencialidad:** La información manejada por el sistema está protegida de acceso y divulgación no autorizada. Esto se deberá lograr mediante el uso e implementación de un control de usuarios.

Usabilidad

- Los elementos de interfaz de usuario tienen que poseer un nombre sugerente de acuerdo a sus funcionalidades para que el usuario la perciba de manera intuitiva. Debe mostrar al menos una breve descripción de cómo se utiliza o para que funciona cada uno.
- Los formularios deben brindar la información que especifique las reglas de los campos que posee como por ejemplo aquellos que sean obligatorios, los de tipo de dato específicos o longitudes de entradas ajustadas a un tamaño específico.
- Los contenidos deben poseer una correcta validez gramatical y ortográfica.

- La presentación de las interfaces visuales al usuario debe hacerse de manera coherente y de forma amigable, permitiendo así una fácil interacción y un rápido entendimiento del sistema por parte de nuevos usuarios.

Software

La característica de la aplicación de ser una plataforma web, brinda al usuario la posibilidad de que con solo un navegador web pueda realizar todas las operaciones posibles en la misma, pero dada las tecnologías utilizadas para su implementación se recomienda utilizar los siguientes.

- Mozilla Firefox
- Chrome
- Opera
- Safari

Hardware

Para el correcto funcionamiento del módulo se requiere para el servidor:

- PC Dual Core
- CPU 1.9 GHz
- 4 GB de RAM
- 500 GB de disco duro

Restricciones de diseño

El sistema deberá poseer una interfaz gráfica uniforme definida por XILEMA como estrategia marcaría, estructurado por módulos que integran la aplicación, reuniendo todas las funcionalidades de la aplicación, a las que se pueden acceder mediante un menú estos módulos y funcionalidades se encuentran en un panel ubicado a la izquierda en cada vista del sistema. La interfaz incluirá solo la información y elementos necesarios que hagan fácil su utilización al usuario.

2.4 Exploración

En la fase de Exploración los clientes escriben las historias de usuario (funcionalidades con que debe contar el sistema). Cada plantilla describe las características que deben ser adicionadas al programa. Al mismo tiempo el equipo del proyecto se familiariza con las herramientas, la tecnología y las prácticas que utilizarán en el proyecto. La tecnología a ser usada será probada y las posibles arquitecturas para el sistema son exploradas construyendo un prototipo del sistema. La fase de exploración toma entre unas cuantas semanas a unos cuantos meses, dependiendo de que tanto los programadores conocen la tecnología. (Beck, 2000)

2.4.1 Historia de usuario (HU)

Las HU son la forma en que se especifican los requisitos del sistema con el uso de la metodología XP. Estas se escriben desde la perspectiva del cliente, aunque los desarrolladores pueden brindar también su ayuda

en la identificación de las mismas.(Beck 2000) Son escritas en lenguaje natural, sin un formato predeterminado, no excediendo su tamaño de unas pocas líneas de texto. Además, guían la construcción de las pruebas de aceptación y son utilizadas para estimar tiempos de desarrollo.

Las HU definidas se clasificaron atendiendo a diferentes parámetros como se muestra en la Tabla 2, facilitando así su organización y la correcta planificación del proyecto. Los parámetros definidos en cada una de las HU se detallan a continuación.

Tabla 2. Parámetros para las HU.

Parámetros	Descripción
Nombre.	Para identificar la HU.
Prioridad.	Esta se clasifica en: <ul style="list-style-type: none"> ➤ Alta: Constituyen funcionalidades principales del sistema, o forman parte esencial de la arquitectura del mismo. ➤ Media: Son funcionalidades importantes y de gran valor para el usuario pero que no impiden poner el proyecto en marcha si no se tienen. ➤ Baja: Son funcionalidades que sería deseable tener y podrían incluirse en caso de que hubiese recursos para ello.
Riesgo de Desarrollo.	Se clasifica en: <ul style="list-style-type: none"> ➤ Alto: En caso de tener algún error de implementación, pueden afectar la disponibilidad del sistema. ➤ Medio: En caso de presentar errores retrasan la entrega de la versión. ➤ Bajo: En caso de presentar errores, estos pueden ser tratados con facilidad y no afectan el desarrollo del proyecto.
Iteración asignada.	Número de la iteración a la que ha sido asignada.
Puntos estimados.	Tiempo estimado de desarrollo para completar la HU. Un punto equivale a una semana ideal de programación (40 horas).
Descripción.	Es donde se define la funcionalidad que se quiere desarrollar.
Observaciones.	Detalles a tener en cuenta para desarrollar la HU correctamente.

Se identificaron 11 HU:

Tabla 3. HU – Autenticar usuario.

Historia de Usuario

Nombre: Autenticar usuario.	
Prioridad: Baja.	Riesgo de desarrollo: Baja.
Iteración Asignada: 5	Puntos Estimados: 0.4
Descripción: Se autentica el usuario en el sistema, quedando registrado como activo.	
Observaciones: El usuario se autenticará en el sistema con un nombre y una contraseña predefinida por el cliente, al realizar esta acción puede acceder a todas las funcionalidades del sistema. Dependiendo de los privilegios que tenga el usuario puede acceder a la administración de usuarios, funcionalidad inicialmente definida para el cliente que tenga el rol de usuario base.	
Prototipo:	
<div style="margin-bottom: 10px;"> <input type="text" value="usuario"/> </div> <div style="margin-bottom: 10px;"> <input type="text" value="contraseña"/> </div> <div style="margin-bottom: 10px;"> <input type="checkbox"/> recordarme </div> <div> <input type="button" value="Entrar"/> </div>	

Tabla 4. HU - Gestionar temáticas.

Historia de Usuario	
Nombre: Gestionar temáticas.	
Prioridad: Alta.	Riesgo de desarrollo: Medio.
Iteración Asignada: 3	Puntos Estimados: 2
Descripción: Se crean perfiles de cada temática asociada a productos del centro para incluir en ellos los selectores de búsqueda dentro de la información adquirida. Estos perfiles permiten al sistema una evaluación de los resultados por tema y la interacción con el usuario a partir de estos resultados para realizar la retroalimentación.	
Observaciones: El usuario puede añadir, modificar, mostrar y eliminar temáticas. Se muestran bloques de información por temática. Se introducen términos lingüísticos que definen estos bloques en clases, según su importancia.	

Prototipo:

Checkbox

Checkbox

Tabla 5. HU - Gestionar de selectores.

Historia de Usuario													
Nombre: Gestionar selectores.													
Prioridad: Media	Riesgo de desarrollo: Baja												
Iteración Asignada: 4	Puntos Estimados: 1												
Descripción: Se crean selectores que representan términos y un peso ligado a ellos, que indica el interés del usuario sobre ese término en el tema al que pertenece.													
Observaciones: El usuario puede añadir, modificar, mostrar, listar y eliminar selectores dentro de los temas a los que pertenezcan y asociarlos a las URLs dentro de cada tema.													
Prototipo:													
<div style="display: flex; flex-direction: column; gap: 10px;"> <input type="button" value="Agregar"/> <input type="button" value="Eliminar"/> <input type="button" value="Modificar"/> </div>	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr style="background-color: #cccccc;"> <th style="width: 20%;">seleccionar</th> <th style="width: 25%;">tema</th> <th style="width: 20%;">palabras clave</th> <th style="width: 35%;">peso</th> </tr> </thead> <tbody> <tr> <td style="text-align: left; vertical-align: middle;">○</td> <td>Tema 1</td> <td>search</td> <td>0.6</td> </tr> <tr> <td> </td> <td> </td> <td> </td> <td> </td> </tr> </tbody> </table>	seleccionar	tema	palabras clave	peso	○	Tema 1	search	0.6				
seleccionar	tema	palabras clave	peso										
○	Tema 1	search	0.6										

Tabla 6. HU - Gestionar URLs.

Historia de Usuario													
Nombre: Gestionar URLs													
Prioridad: Media	Riesgo de desarrollo: Medio												
Iteración Asignada: 4	Puntos Estimados: 1.4												
Descripción: Permite añadir, modificar, mostrar, listar o eliminar las URL a las que se le realizará el monitoreo.													
Observaciones: El sistema muestra sitios web referentes a una búsqueda realizada por el usuario por palabras claves, presentando la posibilidad de seleccionar las URLs de los sitios que se consideren importantes. Permite hacer una inclusión de cada URL a las temáticas que se hallan definido. Las URLs también se pueden incluir de forma manual por el usuario.													
Prototipo:													
<div style="display: flex; justify-content: space-between; align-items: flex-start;"> <div style="display: flex; flex-direction: column; gap: 10px;"> <input type="button" value="Agregar"/> <input type="button" value="Eliminar"/> <input type="button" value="Modificar"/> </div> <table border="1" style="border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="width: 10%;">seleccionar</th> <th style="width: 20%;">Nombre</th> <th style="width: 20%;">Direccion</th> <th style="width: 20%;">tematica</th> </tr> </thead> <tbody> <tr> <td><input type="radio"/></td> <td>Url1</td> <td>http://sffsf.com</td> <td>Tema 1</td> </tr> <tr> <td> </td> <td> </td> <td> </td> <td> </td> </tr> </tbody> </table> <div style="display: flex; justify-content: flex-end; align-items: center;"> <input type="button" value="Iniciar monitoreo"/> </div> </div>		seleccionar	Nombre	Direccion	tematica	<input type="radio"/>	Url1	http://sffsf.com	Tema 1				
seleccionar	Nombre	Direccion	tematica										
<input type="radio"/>	Url1	http://sffsf.com	Tema 1										

Tabla 7. HU - Iniciar y detener el monitoreo de las URLs.

Historia de Usuario	
Nombre: Iniciar y detener el monitoreo de las URLs.	
Prioridad: Alta	Riesgo de desarrollo: Alta
Iteración Asignada: 1	Puntos Estimados: 3
Descripción: Inicia y detiene el monitoreo de las URLs añadidas a cada tema, ejecutando el <i>web scraping</i> como método de recopilación de la información.	
Observaciones: El sistema inicia una búsqueda de información dentro de los sitios de cada URL. Se realiza de forma controlada por un temporizador para evitar bloqueos por parte de los sitios web, ya que se generarían peticiones que no corresponderían con el comportamiento de una persona.	
Prototipo: N.A	

Tabla 8. HU - Configurar automáticamente el monitoreo.

Historia de Usuario													
Nombre: Configurar automáticamente el monitoreo.													
Prioridad: Media	Riesgo de desarrollo: Media												
Iteración Asignada: 3	Puntos Estimados: 1												
Descripción: Se configura el monitoreo de las URLs de forma automática.													
Observaciones: El sistema inicia el monitoreo de forma automática configurada por el usuario en una fecha y hora especificada.													
Prototipo:													
<div style="display: flex; align-items: center; gap: 20px;"> <div style="display: flex; flex-direction: column; gap: 10px;"> <input type="button" value="Agregar"/> <input type="button" value="Eliminar"/> </div> <table border="1"> <thead> <tr> <th>Activar</th> <th>Fecha</th> <th>Dirección</th> <th>Estado</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;"><input checked="" type="checkbox"/></td> <td>14/05/02 10:40 am</td> <td>http://sffsf.com</td> <td>Activa</td> </tr> <tr> <td> </td> <td> </td> <td> </td> <td> </td> </tr> </tbody> </table> </div>		Activar	Fecha	Dirección	Estado	<input checked="" type="checkbox"/>	14/05/02 10:40 am	http://sffsf.com	Activa				
Activar	Fecha	Dirección	Estado										
<input checked="" type="checkbox"/>	14/05/02 10:40 am	http://sffsf.com	Activa										

Tabla 9. HU - Listar el registro de la búsqueda y el número de URLs analizadas.

Historia de Usuario	
Nombre: Listar el registro de la búsqueda y el número de URLs analizadas.	
Prioridad: Baja	Riesgo de desarrollo: Baja
Iteración Asignada: 4	Puntos Estimados: 0.6
Descripción: El sistema muestra las trazas de inicio de búsquedas y la cantidad de URLs analizadas.	
Observaciones: Presenta una interfaz donde se tiene todas las búsquedas anteriores con datos de fecha y hora de inicio y culminación de cada una.	

Prototipo:

Fecha	Cantidad de url Analizadas
14/05/02 10:40 am	3

Tabla 10. HU - Gestionar los resultados de las búsquedas.

Historia de Usuario	
Nombre: Gestionar los resultados de las búsquedas.	
Prioridad: Alta	Riesgo de desarrollo: Alto
Iteración Asignada: 2	Puntos Estimados: 3
Descripción: El sistema es capaz de realizar una búsqueda de los selectores definidos por el usuario en los bloques de información obtenidos, determinando así la importancia de cada bloque por la cantidad de selectores incluidos en él.	
Observaciones: Se utiliza una función que vincule los documentos con cada término de los selectores mediante una comparación booleana de presencia/ausencia dentro del documento, permitiendo que según sus pesos se evalúe el documento según clases lingüísticas definidas.	

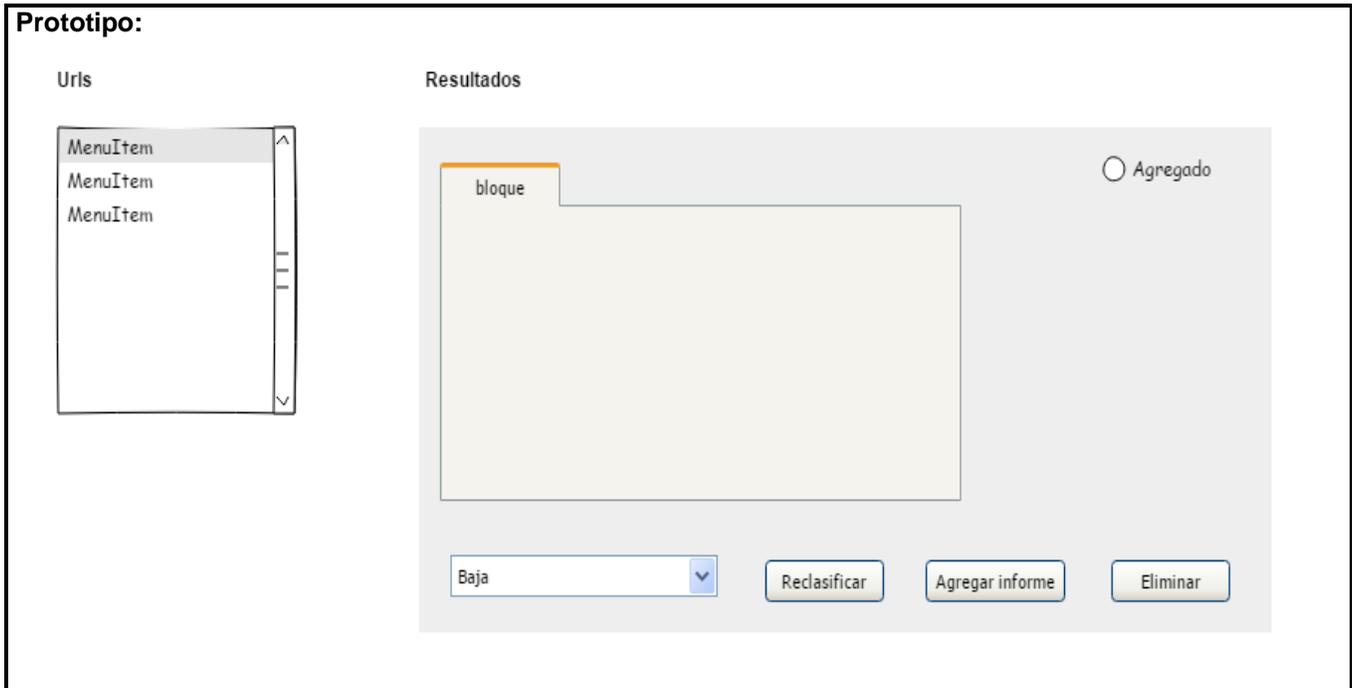


Tabla 11. HU - Generar documento PDF con resultados seleccionados.

Historia de Usuario	
Nombre: Generar documento PDF con resultados seleccionados.	
Prioridad: Baja	Riesgo de desarrollo: Baja
Iteración Asignada: 5	Puntos Estimados: 0.6
Descripción: Se genera un documento, seleccionando los resultados más relevantes para el usuario.	
Observaciones: El sistema genera de forma automática o manual un informe estructurado por temáticas de los resultados obtenidos.	
Prototipo: N.A	

Tabla 12. HU - Configurar envío de informes por correo.

Historia de Usuario	
Nombre: Configurar envío de informes por correo.	
Prioridad: Media	Riesgo de desarrollo: Baja

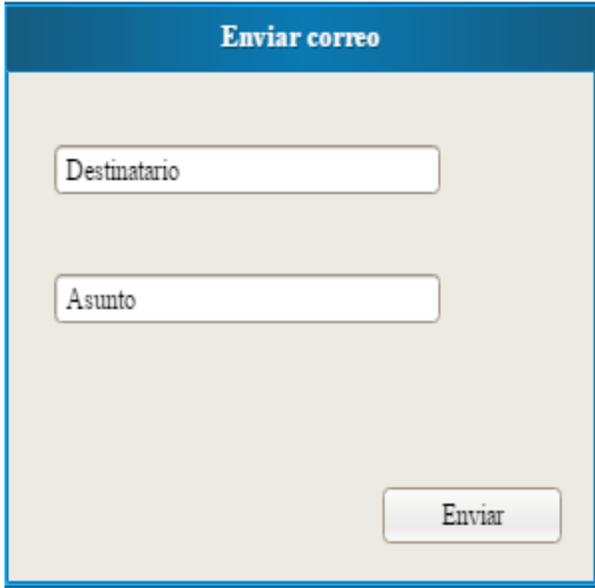
Iteración Asignada: 5	Puntos Estimados: 1
Descripción: Se realiza el envío de los informes generados a una cuenta de correo especificada por el usuario.	
Observaciones: Los informes enviados pueden ser únicos o ser una recopilación de varios realizados por el sistema en un período de tiempo. Pueden ser enviados los informes dentro del cuerpo del correo o ser adjuntado el documento PDF con los resultados del análisis.	
Prototipo:	
	

Tabla 13. HU – Configurar envío de forma automática de los resultados.

Historia de Usuario	
Nombre: Configurar envío de forma automática de los resultados.	
Prioridad: Baja	Riesgo de desarrollo: Baja
Iteración Asignada: 5	Puntos Estimados: 1
Descripción: El sistema es capaz de permitirle al usuario configurar el envío de los correos de manera automática o manual.	
Observaciones: --	

Prototipo:

Envío de resultados Automático

Agregar

Eliminar

Marcado para envío	Fecha	Dirección de correo destinatario	Estado
<input checked="" type="checkbox"/>	20/4/18	cbenitez@estudiantes.uci.cu	Enviada

2.5 Planeación

En esta fase el cliente establece la prioridad de cada historia de usuario, y correspondientemente, los programadores realizan una estimación del esfuerzo necesario de cada una de ellas. Se toman acuerdos sobre el contenido de la primera entrega y se determina un cronograma en conjunto con el cliente. Las estimaciones de esfuerzo asociado a la implementación de las HU la establecen los programadores utilizando como medida el punto. Un punto equivale a una semana ideal de programación o 40 horas. (Beck 2000)

2.5.1 Estimación de esfuerzo por HU

Para el desarrollo de la aplicación propuesta se ha realizado una estimación de esfuerzo por cada una de las Historias de Usuario identificadas, resultados que se muestran a continuación:

Tabla 14. Estimación de esfuerzo por HU.

Historia de Usuario	Puntos de Estimación
Autenticar usuario.	0.4
Gestionar temáticas.	2
Gestionar selectores.	1
Gestionar URLs.	1.4
Iniciar y detener el monitoreo de las URLs.	3
Configurar automáticamente el monitoreo.	1
Listar el registro de la búsqueda y el número de URLs analizadas.	0.6
Gestionar los resultados de las búsquedas.	3

Generar documento PDF con resultados seleccionados.	0.6
Configurar envío de informes por correo.	1
Configurar envío de forma automática de los resultados.	1

2.5.2 Plan de iteraciones

Las historias de usuarios seleccionadas para cada entrega son desarrolladas y probadas en un ciclo de iteración, de acuerdo al orden preestablecido. Sobre la base de lo antes mencionado se decide realizar esta en 5 iteraciones, las cuales se detallan a continuación.

Iteración #1: Esta iteración tiene como objetivo la implementación de la historia de usuario Iniciar y detener el monitoreo de las URLs.

Iteración #2: El objetivo de esta iteración es la implementación de las HU Gestionar los resultados de las búsquedas.

Iteración #3: Durante el transcurso de esta iteración se implementan las HU Gestionar temáticas y Configurar automáticamente el monitoreo.

Iteración #4: Se implementará en esta iteración las HU Gestionar selectores, Gestionar URLs y Listar el registro de la búsqueda y el número de URLs analizadas.

Iteración #5: En esta iteración se tiene como objetivo la implementación de las HU Autenticar usuario, Generar documento PDF con resultados seleccionados, Configurar envío de informes por correo y Configurar envío de informes por correo.

2.5.3 Plan de duración de las iteraciones

Como parte del ciclo de vida de un proyecto que utiliza la metodología XP se crea el plan de duración de cada una de las iteraciones. Este plan se encarga de mostrar las HU que serán afrontadas en cada una de las iteraciones, así como la duración estimada de estas últimas y el orden en que se implementarán.

Tabla 15. Plan de duración de las iteraciones.

Iteración	Orden de la HU a Implementar	Duración Total (Semanas)
Iteración 1	Iniciar y detener el monitoreo de las URL	3
Iteración 2	Gestionar los resultados de las búsquedas	3
Iteración 3	Gestionar temáticas	2
	Configurar automáticamente el monitoreo	1

Iteración 4	Gestionar selectores	1
	Gestionar URLs	1.4
	Listar el registro de la búsqueda y el número de URLs analizadas.	0.6
Iteración 5	Generar documento PDF con resultados seleccionados	0.6
	Autenticar usuario.	0.4
	Configurar envío de informes por correo	1
	Configurar envío de forma automática de los resultados.	1

2.5.4 Plan de entrega

Establece que las historias de usuarios serán agrupadas para conformar una entrega y el orden de las mismas. Este cronograma será el resultado de una reunión entre todos los actores del proyecto. (2016)

A continuación, se presenta el plan de entregas elaborado para la fase de implementación, que comienza el lunes 7 de febrero de 2018. Como producto del mismo se harán entregas incrementales del sistema al finalizar cada iteración, en la fecha aproximada que se indica en la siguiente tabla.

Tabla 16. Plan de entrega.

Iteración	Fecha de Entrega
Iteración 1	27 de febrero del 2018
Iteración 2	20 de marzo del 2018
Iteración 3	10 de abril del 2018
Iteración 4	1 de mayo del 2018
Iteración 5	22 de mayo del 2018

2.6 Diseño

Durante el diseño de la solución, la máxima simplicidad posible es la clave para el éxito de XP. Se debe diseñar la solución más simple que pueda funcionar y ser implementada en un momento determinado del proyecto. La complejidad innecesaria y el código extra deben ser evitados en todo momento. (Cockburn 2000)

El diseño adecuado para el software es aquel que supera con éxito todas las pruebas, refleja claramente a intención de implementación de los programadores y tiene el menor número posible de clases y métodos. Se debe tener en cuenta que un diseño complejo siempre tarda más en desarrollarse que uno simple, y que siempre es más fácil añadir complejidad a un diseño simple que quitarla de uno complejo. (Pressman 2005)

2.6.1 Patrón y estilo de arquitectura.

La arquitectura del software alude a “la estructura general de éste y a las formas en las que ésta da integridad conceptual a un sistema.” (Shaw, y otros, 1995) En su forma más sencilla, la arquitectura es la estructura de organización de los componentes de un programa (módulos), la forma en la que éstos interactúan y la estructura de datos que utilizan. Sin embargo, en un sentido más amplio, los componentes se generalizan para que representen los elementos de un sistema grande y sus interacciones. (Pressman, 2005)

Modelo-Plantilla-Vista (MTV)

Django, el framework de desarrollo seleccionado del lado del servidor, respeta el paradigma conocido como Modelo Plantilla Vista (MTV), el cual está fuertemente inspirado en la filosofía de desarrollo Modelo Vista Controlador (MVC), con ciertas diferencias. Como resultado, lo que se llamaría controlador en un verdadero framework MVC se llama en Django vista, y lo que se llamaría vista se llama plantilla. Para un mejor entendimiento se presenta el siguiente diagrama:

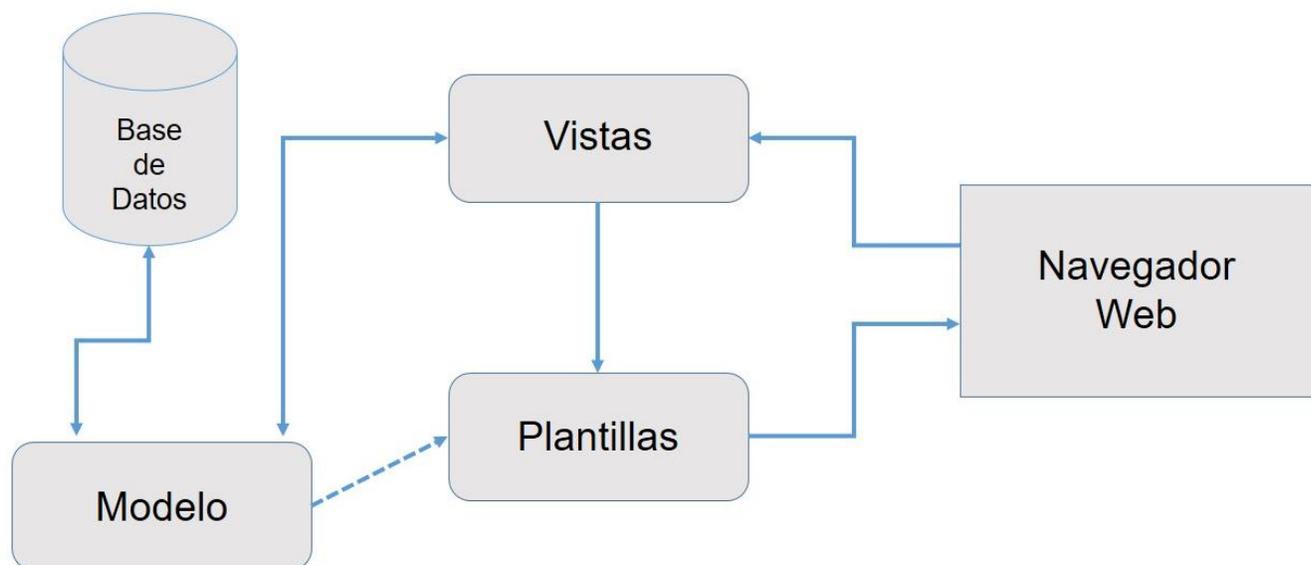


Fig. 5 Diagrama del patrón Modelo Plantilla Vista (MTV)

El patrón MTV al igual que el MVC permite separar los datos, la lógica de la aplicación y la presentación gráfica con la finalidad de conseguir un código limpio y un fácil mantenimiento ya que las modificaciones de una parte de la aplicación no afectan a las demás. (Django Software Foundation, 2005-2018)

2.6.2 Patrones de diseño

Un patrón de diseño describe una estructura de diseño que resuelve un problema particular del diseño dentro de un contexto específico y entre medios que afectan la manera en la que se aplica y en la que se utiliza dicho patrón.

El objetivo de cada patrón de diseño es proporcionar una descripción que permita a un diseñador determinar

- 1) Si el patrón es aplicable al trabajo en cuestión,
- 2) si puede volverse a usar (con lo que se ahorra tiempo de diseño) y;
- 3) si sirve como guía para desarrollar un patrón distinto en funciones o estructura. (Pressman, 2005)

Para el diseño de la aplicación se hizo uso de los Patrones Generales de Software para Asignar Responsabilidades (GRASP), así como de los patrones Gang of Four o Pandilla de los Cuatro (GoF).

Patrones GRASP

Los patrones GRASP describen los principios fundamentales del diseño de objetos y la asignación de responsabilidades (Craig 2001), en el sistema se encuentran los siguientes:

- **Alta cohesión:** Aplicado a las clases del diseño, ya que en cada una solo se implementan las funcionalidades que le corresponden, la clase `models` por ejemplo, solo implementa la instanciación de objetos ORM definidos por Django para trabajar los modelos de la base de datos, como se muestra a continuación en el código:

```
from django.db import models

# Create your models here.
class Tematica(models.Model):
    name = models.CharField(max_length=200)
    descripcion = models.TextField(default='')
    fecha_inclusion = models.DateTimeField(auto_now_add=True)
```

Fig. 6 Evidencia del uso del patrón Alta cohesión

- **Bajo acoplamiento:** La comunicación con otras clases es escasa, ya que cada clase interactúa con el menor número posible de clases. Esta interacción se evidencia en todas las vistas como clases controladoras del sistema como se puede observar en la siguiente Fig. 7:

```

@login_required(login_url='/login/')
def search_google(request):
    return render(request, 'busqueda.html')

def search_google_post(request):
    if request.is_ajax and request.method == 'POST':
        var = request.POST['enviado']
        temas = Tematica.objects.all()
        lista_url = {'lista_url': scrap(var), 'lista_tema':
temas}
    return render(request, 'tabla_busqueda.html', lista_url)

```

Fig. 7 Evidencia del uso del patrón Bajo acoplamiento

- **Creador:** Las clases que crean objetos como funcionalidad contienen toda la información necesaria para construir los mismos. Este patrón se evidencia en la clase que realiza el *web scraping* ya que instancia un objeto con toda la información obtenida por la misma, mostrándose a continuación:

```

class Ob:
    url = ''
    tema = ''
    def __init__(self, url, tema):
        self.url = url
        self.tema = tema
class selector:
    id = 0
    nuevo = 0
    def __init__(self, id, nuevo):
        self.id = id
        self.nuevo = nuevo

```

Fig. 8 Evidencia del uso del patrón Creador

- **Experto:** Las tareas realizadas por los objetos se hacen con la propia información incluida en ellos. Se distribuye el comportamiento entre las clases que contienen la información requerida. Son más fáciles de entender y mantener. Este patrón se representa en todas las vistas codificadas en la aplicación. Un ejemplo de su uso se muestra en la siguiente Fig. 9:

```

def scrap_feed_rss(urls):
    string_file = ""
    list_document = []
    doc = document
    if hasattr(ssl, '_create_unverified_context'):
        ssl._create_default_https_context =
ssl._create_unverified_context
        for dir in urls:
            rss = feedparser.parse(dir.url, handlers=[proxy])
            print rss
            # se recorre todos los post que aparecen en el RSS
            # para luego mostrar solamente el titulo, el link y
la descripcion
            for post in rss.entries:
                # string_file += post.description
                list_document.append(doc(dir.tema, post.title,
post.link, dir.url, strip_tags(post.description)))
                # string_file += '\n'
                # print post.title + ":" + post.link+"
"+post.description
                # imprimir_file(string_file.encode('utf-8'))
    return list_document

```

Fig. 9 Evidencia del uso del patrón Experto

Patrones GoF

Los patrones de diseño GoF “describen soluciones simples y elegantes a problemas específicos en el diseño de software orientado a objetos”.(Craig 2001) Los utilizados en el diseño de la solución fueron:

- **Front Controller (Controlador Frontal):** Django posee una implementación de Controlador Frontal que despacha las peticiones hacia métodos o clases, que en la práctica son páginas controladoras. Antes del despacho, la petición es procesada por varios filtros (middlewares).
- **Bridge:** El objetivo de este patrón es la separación (desacoplamiento) de una abstracción de su implementación, de tal forma que ambas puedan ser modificadas independientemente sin afectar la una a la otra. Se evidencia en las diferentes implementaciones que se le realizan a la clase scrap, ya que las subclases que heredan de ella nos permiten hacer obtención de información de la misma forma, pero dando como resultado instancias diferentes para cada una. A continuación, se muestra un ejemplo de su uso:

```

class document:
    tema = ''
    title = ''
    link = ''
    urlbase = ''
    descripcion = ''

    def __init__(self, tema, title, link, urlbase, descripcion):
        self.link = link
        self.urlbase = urlbase
        self.tema = tema
        self.title = title
        self.descripcion = descripcion

class rss_directo_auto(document):
    def scrap_feed_auto(url, tematica):
        list_document = []
        doc = document

class rss_directo(document):
    def scrap_feed_rss(urls):
        string_file = ""
        list_document = []
        doc = document

```

Fig. 10 Evidencia del uso del patrón Bridge

- **Singleton (Patrón Solitario):** Este patrón se asegura de que de una clase solo exista una instancia y que se ofrezca un punto de acceso a ella. Se emplea en la clase ProxiDir la cual brinda la configuración para acceder a un proxy que permita la conexión con internet, como se muestra en el código de la Fig. 11:

```

class FeedFinder:
    # metodo que realiza la petición a la pagina y devuelve su
    html
    def geturl(self, urlBase):
        proxyDict = {
            'http': '127.0.0.1:3129',
            'https': '127.0.0.1:3129'
        }
        # ie=UTF-8&oe=UTF-8&
        # Realizamos la petición a la web
        html = requests.get(urlBase, proxies=proxyDict,
            verify=False)
        return html

```

Fig. 11 Evidencia del uso del patrón Singleton

2.6.3 Tarjetas CRC

Las tarjetas CRC (Clase Responsabilidad- Colaborador) son confeccionadas durante la etapa de diseño de la metodología de desarrollo XP. La misma constituye un modelo simple que tiene como objetivo desarrollar una representación organizada de las clases. Cada tarjeta representa una clase, de cada una se describe las responsabilidades que tiene y las clases colaboradoras que se relacionan con la misma. Una responsabilidad es cualquier cosa que la clase sabe o hace. Los colaboradores son aquellas clases que se requieren para que una clase reciba la información necesaria para completar una responsabilidad. (Beck, 2000)

A continuación, se describen las tarjetas CRC obtenidas durante la etapa de Diseño del presente trabajo:

Tabla 17. Tarjeta CRC #1

Tarjeta CRC #1	
Clase: Models	
Responsabilidad	Colaboración
<ul style="list-style-type: none"> ➤ Guarda y recupera desde la base de datos la información asociada a cada uno de los modelos del sistema. ➤ Define el modelo de datos a través del ORM de django. 	.django.db.models

Tabla 18. Tarjeta CRC #2

Tarjeta CRC # 2	
Clase: Views	
Responsabilidad	Colaboración
<ul style="list-style-type: none"> ➤ Provee las APIS correspondientes al sistema de autenticación y los modelos del sistema. ➤ Gestiona las peticiones que se le realizan al servidor, GET, POST, DELETE y OPTION desde las APIS generadas. 	<ul style="list-style-type: none"> ➤ Models ➤ Search

Tabla 19. Tarjeta CRC #3

Tarjeta CRC # 3
Clase: Search

Responsabilidad	Colaboración
<ul style="list-style-type: none"> ➤ Posibilita la conexión y petición de consultas a las páginas web y al proxy seleccionado. ➤ Extrae la información que será procesada desde las páginas web de interés del usuario. 	<ul style="list-style-type: none"> ➤ Request ➤ BeautifullSoup ➤ View

Tabla 20. Tarjeta CRC #4

Tarjeta CRC # 4	
Clase: search_extern	
Responsabilidad	Colaboración
<ul style="list-style-type: none"> ➤ Define los atributos para un objeto utilizado en <i>search</i>. 	<ul style="list-style-type: none"> ➤ Search

Tabla 21. Tarjeta CRC #5

Tarjeta CRC # 5	
Clase: scrap_feed	
Responsabilidad	Colaboración
<ul style="list-style-type: none"> ➤ Permite la obtención de la estructura de los RSS desde una o varias URLs obtenidas desde ob. ➤ Realiza la instanciación de los datos que sean brindados por los RSS. ➤ Devuelve una lista de objetos del tipo <i>document</i> a <i>views</i>. 	<ul style="list-style-type: none"> ➤ Feedparser ➤ ssl ➤ View ➤ document ➤ ob

Tabla 22. Tarjeta CRC # 6

Tarjeta CRC # 6	
Clase: feed_finder	
Responsabilidad	Colaboración
<ul style="list-style-type: none"> ➤ Realiza una petición a una URL seleccionada, obteniendo como resultado el HTML de la página. ➤ Extrae los enlaces RSS que contenga el HTML extraído desde la URL. 	<ul style="list-style-type: none"> ➤ scrap_feed ➤ beautifullSoup ➤ feedparser

Tabla 23. Tarjeta CRC # 7

Tarjeta CRC # 7	
Clase: Ob	
Responsabilidad	Colaboración
<ul style="list-style-type: none"> ➤ Define los atributos para un objeto utilizado en scrap_feed. ➤ Contiene la información de las URLs seleccionadas para ser analizadas, además del tema al que pertenecen cada una. 	<ul style="list-style-type: none"> ➤ scrap_feed ➤ view

Tabla 24. Tarjeta CRC # 8

Tarjeta CRC # 8	
Clase: document	
Responsabilidad	Colaboración
<ul style="list-style-type: none"> ➤ Define los atributos para un objeto utilizado en scrap_feed. ➤ Contiene la información que es obtenida mediante la clase scrap_feed, creando un objeto con esta información. 	<ul style="list-style-type: none"> ➤ scrap_feed ➤ view

2.7 Diseño de la Base de Datos

En cualquier sistema en el que se gestione información, la base de datos desempeña un papel fundamental. El diseño de la misma es un proceso esencial para el desarrollo de sistemas de este tipo, pues la misma permite el almacenamiento de la información de forma coherente y organizada, evitando así pérdidas e inconsistencias, de esta manera la recuperación de la información se puede realizar de forma rápida y flexible. A continuación, se muestra el modelo físico de la base de datos que se utilizó:



Fig. 12 Modelo físico de la base de datos

2.8 Algoritmo a desarrollar

Para dar solución al problema de investigación se implementó un SRI, que se encarga de obtener datos desde los diferentes sitios web seleccionados por el usuario, teniendo como deficiencia la incapacidad de clasificar los resultados, lo que aumenta la posibilidad de intoxicación del usuario.

Para evitar la sobreinformación se utilizó un algoritmo que emplea la lógica difusa con retroalimentación aplicada a los SRIs, lo que facilita la interacción con el usuario. Inicialmente el usuario define los selectores (términos) y les asigna un peso inicial, que representa el grado de importancia de ese término dentro de la información que quiere obtener. Este algoritmo busca selectores dentro de los bloques de información que se obtienen con el SRI desde un sitio web. Luego, analiza cada uno de los bloques de información, aplicando una función de fuzificación que promedia los pesos de todos los términos encontrados y emite una clasificación que puede ser buena, regular o mala para cada bloque. Esta clasificación dada por el sistema, puede ser reevaluada por el usuario en caso de que los términos de la consulta no arrojen los resultados esperados. De esta nueva evaluación el sistema aprende que, para los bloques de información obtenidos

con el SRI en el sitio web, los pesos de relevancia son distintos, creando un perfil de usuario con los selectores que se analizaron en ese bloque de información y sus pesos correspondientes, la dirección URL de donde se obtuvo la información y un nuevo peso asociado a cada selector, que le permita reconocer ante una nueva búsqueda los selectores de mayor importancia dentro de ese sitio web específico. Debe quedar claro que la clasificación de la información, siempre queda en manos del usuario, dado que el sistema aprende de sus decisiones siendo estas correctas o no, por lo que se recomienda el juicio de un experto para la reclasificación y análisis de la información obtenida.

Seudocódigo del algoritmo a implementar:

Lista1 = Obtener bloques (*términos*, *URLs*)

Para cada elemento de *Lista1*:

Lista2 = Encontrar selectores (*términos*)

Para cada URL

Encontrar URL en perfil de usuario

Si está en perfil:

Peso = peso nuevo en perfil para elemento actual de *Lista2*

Si no está entonces:

Peso += peso de elemento actual de *Lista2*

Resultado = promedio entre peso y total de elementos de *Lista2*

Clasificar elemento *Lista1* (*Resultado*):

Si *resultado* > 0 y *resultado* < 0.3

Entonces *clasificación* = baja

Si *resultado* >= 0.3 y *resultado* < 0.6

Entonces *clasificación* = media

Si *resultado* >= 0.6 y *resultado* <= 1

Entonces *clasificación* = alta

Salida = bloque clasificado

Obtener bloque (*URLs*, *términos*):

Cambio de clasificación (*nueva*, *anterior*):

Si *nueva*=mala y *anterior*=media

Resultado = -1

Si *nueva* = mala y *anterior* = buena

Resultado = -2

Si nueva = media y anterior= mala

Resultado = 1

Si nueva = media y anterior= buena

Resultado = -1

Si nueva = buena y anterior= mala

Resultado = 2

Si nueva = buena y anterior= media

Resultado = 1

Salida = *Resultado*

Reclasificar bloque (*nueva clasificación, clasificación anterior, peso*):

*Peso nuevo = peso + 0.2 * cambio de clasificación (nueva clasificación, anterior clasificación)*

Crear perfil (*peso nuevo, URL, selectores*)

Salida = bloque reclasificado, nuevo perfil

- Para cada término dentro de un bloque de información sumar peso.
- Promediar con total de términos empleados para obtener todos los bloques de información.
- El resultado se utiliza para clasificar el bloque con términos lingüísticos de bueno, regular o malo.
- Se presenta el resultado al cliente.
- El resultado puede ser reclasificado según las necesidades del cliente.
- Si es reclasificado se calculan nuevos pesos para utilizar en nuevas consultas realizadas por el usuario, haciendo una variación de 0.2 según clasificación anterior.
- Si el cliente presenta mejor clasificación que la actual +0.2 al peso del término.
- Si el cliente presenta peor clasificación que la actual -0.2 al peso del término.
- Se crea un perfil de usuario donde se guardan las direcciones de los sitios que fueron consultados y los términos empleados para clasificar la información que desde ellos se obtuvo con sus nuevos pesos.
- Estos perfiles de usuario se utilizarán en las nuevas consultas para clasificar los bloques de información.

2.9 Conclusiones del capítulo

Se elaboraron los artefactos correspondientes a las etapas de Exploración, Planificación y Diseño, definiéndose además las funcionalidades que tendrá el sistema. Se identificaron 11 HU; de ellas tres fueron



catalogadas de prioridad alta, cuatro de prioridad media y cuatro de prioridad baja. Se calculó la planificación del esfuerzo dedicado a la realización de cada una de estas HU en el orden en que se les dará cumplimiento según las necesidades del cliente. Además, permitió definir una fecha estimada en la que se le entregará una primera versión del producto al cliente. Se definieron los patrones arquitectónicos y de diseño usados con el objetivo de lograr una mayor organización en los elementos que conforman la aplicación, así como las tarjetas CRC para una mejor comprensión de la misma. Asimismo, se diseñó la base de datos y se describió el algoritmo a desarrollar para un mejor entendimiento.



CAPÍTULO 3: IMPLEMENTACIÓN Y PRUEBAS.

La metodología XP plantea que el proceso de desarrollo del software se debe realizar de forma iterativa, obteniendo al finalizar de cada iteración un producto funcional, que debe ser probado y mostrado al cliente para así tener una retroalimentación entre la opinión de este y los desarrolladores. En el presente capítulo se describen los estándares de codificación utilizados para obtener un código limpio y legible. Se detallan además las cinco iteraciones llevadas a cabo durante la etapa de construcción del sistema, exponiendo las tareas generadas por cada HU, así como las pruebas unitarias y de aceptación efectuadas sobre la aplicación. Además, se hace la comprobación del funcionamiento del algoritmo de lógica difusa utilizado en la aplicación.

3.1 Estándares de codificación

El término **estándar** tiene su origen etimológico en el vocablo inglés *standard*. El concepto se utiliza para nombrar a aquello que puede tomarse como referencia, patrón o modelo. (Definición.Es, 2008-2018)

La **codificación** es el acto y el resultado de codificar. Este verbo, por su parte, puede aludir a modificar la expresión de un mensaje o a registrar algo a través de las reglas de un código. (Definición.Es, 2008-2018)

De acuerdo a los anteriores conceptos, los estándares de codificación son modelos o patrones codificados a través de las reglas de un código específico. La metodología XP enfatiza la comunicación de los programadores a través del código, por lo cual es indispensable que se sigan ciertos estándares de programación (del equipo, de la organización u otros estándares reconocidos para los lenguajes de programación utilizados).

Para lograr una mejor legibilidad y solidez del código se utiliza la Guía de estilo para el código Python (PEP 8). Esta guía posee una gran variedad de conformidades, entre ellas están:

- Usa 4 espacios por nivel de indentación.
- Limita todas las líneas a un máximo de 120 caracteres.
- Separa funciones de alto nivel y definiciones de clase con dos líneas en blanco; y las definiciones de métodos dentro de una clase son separadas por una línea en blanco.
- Las importaciones deben estar en líneas separadas.
- Utilizar el estilo *UPPER_CASE_WITH_UNDERSCORES* para nombrar clases, y el *lower_case_with_underscores* para funciones y métodos. (Python Software Foundation (US), 2001-2018)



3.2 Iteraciones

En el transcurso de las iteraciones se realiza la implementación de las historias de usuario seleccionadas para cada una de estas y se llevan a cabo las pruebas unitarias y de aceptación asociadas a cada una de las mismas. Al iniciar cada iteración, se lleva a cabo una revisión del plan de iteraciones y se modifica de ser necesario. Como parte de este plan, se descomponen las HU en tareas de ingeniería. Estas tareas son para el uso de los programadores, pueden escribirse utilizando un lenguaje técnico y no necesariamente deben ser entendibles para el cliente. (Beck, 2000)

Siguiendo la planificación confeccionada, se llevaron a cabo cinco iteraciones de desarrollo sobre el sistema, alcanzándose al terminar un producto listo para su despliegue. A continuación, se detallan cada una de las iteraciones.

3.2.1 Pruebas

La fase de las pruebas es una de las etapas fundamentales del desarrollo de una aplicación. El objetivo de cada una de las pruebas no es el de prevenir errores, sino de detectarlos basándose en técnicas y estrategias empleadas en cada una de las pruebas. (Beck, 2000) La metodología XP propone para validar las necesidades de los usuarios y dirigir la implementación las pruebas unitarias y las de aceptación.

Como ya se ha señalado anteriormente existen disimiles estrategias de pruebas, el estudio de este trabajo se ha centrado específicamente en la investigación de las pruebas correspondientes a la metodología de desarrollo de software empleada en este trabajo de diploma. XP divide las pruebas del sistema en dos grupos: pruebas unitarias, encargadas de verificar el código y diseñada por los programadores y las pruebas de aceptación o pruebas funcionales destinadas a evaluar si al final de una iteración se consiguió la funcionalidad requerida diseñadas por el cliente final.

Las **pruebas unitarias** son una de las piedras angulares de XP. Todos los módulos deben de pasar las pruebas unitarias antes de ser liberados o publicados, esto habilita que funcione la propiedad colectiva del código. En este sentido, el sistema y el conjunto de pruebas debe ser guardado junto con el código, para que pueda ser utilizado por otros desarrolladores, en caso de tener que corregir, cambiar o recodificar parte del mismo. (Joskowicz 2008) Para la comprobación del código de la aplicación se utilizó el marco de prueba de Python, denominado "PyUnit" que contiene la clase TestCase donde los casos de prueba y los conceptos de accesorio de prueba son realizados. Cada instancia de TestCase solo se usará para ejecutar un único método de prueba, por lo que se crea un nuevo método para cada prueba. (Python Software Foundation, 1990-2018)



Las **pruebas de aceptación** son creadas en base a las historias de usuarios, en cada ciclo de la iteración del desarrollo. El cliente debe especificar uno o diversos escenarios para comprobar que una historia de usuario ha sido correctamente implementada. Los clientes son responsables de verificar que los resultados de estas pruebas sean correctos. Asimismo, en caso de que fallen varias pruebas, deben indicar el orden de prioridad de resolución. Una historia de usuario no se puede considerar terminada hasta tanto pase correctamente todas las pruebas de aceptación. (Joskowicz, 2008)

Las anteriores pruebas fueron evaluadas automáticamente en cada una de las iteraciones, obteniéndose un resultado final que se muestra en las próximas secciones.

3.3 Iteración 1

En esta iteración se implementó de la historia de usuario Iniciar y detener el monitoreo de las URL.

- HU- Iniciar y detener el monitoreo de las URL.

Tabla 25. Tarea de Ingeniería - Iniciar el monitoreo de las URL.

Tarea de Ingeniería	
Nombre: Iniciar el monitoreo de las URL.	
HU: Iniciar y detener el monitoreo de las URL.	Puntos Estimados: 1.4
Fecha de inicio: 7 de febrero del 2018	Fecha fin: 15 de febrero del 2018
<p>Descripción: El usuario accede en el módulo URLs cargando la vista URLs, la cual renderiza la plantilla correspondiente a las URL almacenadas en la base de datos, donde se posibilita la selección de las direcciones que serán monitoreadas. Se inicia el monitoreo haciendo clic en iniciar monitoreo, esta acción implementa la vista monitorear_ URL la cual recibe desde la plantilla todas las direcciones a monitorear con un objeto json que las contiene. Este objeto accede como parámetro a la clase <i>Feedfinder</i> la cual mediante las librerías <i>feedparser</i> y <i>beautiful soup</i> realizan la selección de la información desde las páginas web.</p>	

Tabla 26. Tarea de Ingeniería - Detener el monitoreo de las URL.

Tarea de Ingeniería	
Nombre: Detener el monitoreo de las URL.	
HU: Iniciar y detener el monitoreo de las URL.	Puntos Estimados: 1.6
Fecha de inicio: 16 de febrero del 2018	Fecha fin: 27 de febrero del 2018
<p>Descripción: El sistema luego de recopilar toda la información desde la página web, termina la conexión con la misma, enviando el resultado de la búsqueda a la clase <i>ReclInfo</i> para ser analizados.</p>	



Pruebas unitarias

En esta iteración se realizaron 4 pruebas unitarias al sistema. Los resultados pueden observarse a continuación:

```

Creating test database for alias 'default'...
System check identified no issues (0 silenced).
....
-----
Ran 4 tests in 0.096s

OK
Destroying test database for alias 'default'...
    
```

Fig. 13 Resultados de las pruebas unitarias de la primera iteración

Pruebas de aceptación

Para realizar las pruebas de aceptación correctamente se elaboraron casos de prueba para cada una de ellas. A continuación, se describe el caso de prueba correspondiente a la HU Iniciar y detener el monitoreo de las URL donde se crearon dos tablas. Las tablas con las pruebas de las iteraciones siguientes deben consultarse en los **Anexos**.

Tabla 27. Caso de Aceptación - Iniciar y detener el monitoreo de las URL - Datos Válidos.

Caso de prueba de aceptación	
Nombre: Iniciar y detener el monitoreo de las URLs - Datos Válidos	
Historia de Usuario: Iniciar y detener el monitoreo de las URLs.	
Descripción: Se prueba el inicio y detención del monitoreo de la URLs guardadas por temáticas.	
Precondiciones:	
<ul style="list-style-type: none"> ➤ Tener al menos una URL agregada. ➤ Autenticarse en el sistema. 	
Pasos de ejecución: Acceder al módulo URLs. Seleccionar las URLs a monitorear.	El sistema muestra una tabla con la lista de URLs. El sistema realiza el monitoreo de las URL seleccionadas, y obtiene toda la información que



Seleccionar el botón Iniciar monitoreo.	brindan los sitios asociados a estas. Durante este proceso muestra una interfaz de espera. Muestra el mensaje de confirmación “Búsqueda realizada, consulte los resultados en el módulo Temas”.
---	--

Tabla 28. Caso de Aceptación - Iniciar y detener el monitoreo de las URL - Datos Inválidos

Caso de prueba de aceptación	
Nombre: Iniciar y detener el monitoreo de las URLs - Datos Inválidos	
Historia de Usuario: Iniciar y detener el monitoreo de las URLs.	
Descripción: Se prueba que el inicio y detención del monitoreo de la URLs guardadas por temáticas sea de manera incorrecta.	
Precondiciones: Autenticarse en el sistema.	
Pasos de ejecución: Acceder al módulo URLs. No se selecciona ninguna URL a monitorear. Seleccionar el botón Iniciar monitoreo.	El sistema muestra una tabla con la lista de URLs. El sistema muestra una alerta: "No hay urls seleccionadas".

3.4 Iteración 2

En el transcurso de esta iteración se implementó la HU Gestionar los resultados de las búsquedas.

- HU-Gestionar los resultados de las búsquedas.

Tabla 29. Tarea de Ingeniería - Añadir los resultados de las búsquedas.

Tarea de Ingeniería	
Nombre: Añadir los resultados de las búsquedas.	
HU: Gestionar los resultados de las búsquedas.	Puntos Estimados: 1
Fecha de inicio: 28 de febrero del 2018	Fecha fin: 6 de marzo del 2018
Descripción: El sistema analiza los bloques de información obtenidos de las URLs monitoreadas a través de la clase ReclInfo, la cual implementa la lógica difusa, esta obtiene desde la base de datos los distintos selectores	



que se le aplicarán a la información, de acuerdo con el tema al que pertenece y aplicando un algoritmo booleano retorna una respuesta fuzificada para cada bloque de información analizado, los mismos quedan clasificados en buena, regular o mala calidad de su información.

Tabla 30. Tarea de Ingeniería - Listar los resultados de las búsquedas.

Tarea de Ingeniería	
Nombre: Listar los resultados de las búsquedas.	
HU: Gestionar los resultados de las búsquedas.	Puntos Estimados: 1
Fecha de inicio: 7 de marzo del 2018	Fecha fin: 13 de marzo del 2018
Descripción: La información ya clasificada, es enviada a la vista mostrar_result con los parámetros definidos por la información, clasificación y temática a la que pertenece. Estos resultados se agrupan según la temática de acuerdo a su clasificación, quedando listados por bloques similares con misma clasificación.	

Tabla 31. Tarea de Ingeniería - Mostrar los resultados de las búsquedas.

Tarea de Ingeniería	
Nombre: Mostrar los resultados de las búsquedas.	
HU: Gestionar los resultados de las búsquedas.	Puntos Estimados: 0.6
Fecha de inicio: 14 de marzo del 2018	Fecha fin: 16 de marzo del 2018
Descripción: El sistema a través de la vista mostrar_result, envía los datos a la plantilla correspondiente a las temáticas con toda la información de las búsquedas realizadas de esa temática, renderizando las mismas con los resultados ya listados.	

Tabla 32. Tarea de Ingeniería - Eliminar los resultados de las búsquedas.

Tarea de Ingeniería	
Nombre: Eliminar los resultados de las búsquedas.	
HU: Gestionar los resultados de las búsquedas.	Puntos Estimados: 0.4
Fecha de inicio: 19 de marzo del 2018	Fecha fin: 20 de marzo del 2018
Descripción: El usuario puede eliminar la información de las búsquedas realizadas, esta acción ejecuta la vista elim_info, la cual elimina la información de las búsquedas realizadas por temáticas, debe confirmarse debido a que es irreversible.	

Pruebas unitarias



En esta iteración se realizaron 13 pruebas unitarias al sistema. Los resultados pueden observarse en la siguiente figura:

```

Creating test database for alias 'default'...
System check identified no issues (0 silenced).
.....
-----
Ran 13 tests in 0.247s

OK
Destroying test database for alias 'default'...
    
```

Fig. 14 Resultados de las pruebas unitarias de la segunda iteración

3.5 Iteración 3

Durante el transcurso de esta iteración se abordó las HU Gestión de temáticas y Configurar automáticamente el monitoreo.

- HU-Gestionar temáticas.

Tabla 33. Tarea de Ingeniería – Modificar temáticas.

Tarea de Ingeniería	
Nombre: Modificar temáticas.	
HU: Gestionar temáticas.	Puntos Estimados: 0.2
Fecha de inicio: 21 de marzo del 2018	Fecha fin: 21 de marzo del 2018
Descripción: El usuario accediendo al módulo Temáticas puede seleccionar la opción modificar temática de las temáticas listadas, donde haciendo clic sobre ella muestra un formulario con los campos de entrada para modificar la temática, al aceptar se ejecuta la vista update_tematica la cual recibe como identificador de la temática seleccionada y sus nuevos datos modificándose en la base de datos.	

Tabla 34. Tarea de Ingeniería - Añadir temáticas.

Tarea de Ingeniería
Nombre: Añadir temáticas.



HU: Gestionar temáticas.	Puntos Estimados: 0.4
Fecha de inicio: 22 de marzo del 2018	Fecha fin: 23 de marzo del 2018
Descripción: El usuario accede al módulo Temáticas, donde están mostradas todas las temáticas que posee registradas el sistema, haciendo clic en agregar temática, se muestra un formulario donde se pasan por parámetros los datos del nombre de la temática, y una breve descripción de la misma, al aceptar los datos entrados se ejecuta la vista guardar_tema quedando agregada la nueva temática en el sistema.	

Tabla 35. Tarea de Ingeniería - Mostrar temáticas.

Tarea de Ingeniería	
Nombre: Mostrar temáticas.	
HU: Gestionar temáticas.	Puntos Estimados: 0.6
Fecha de inicio: 26 de marzo del 2018	Fecha fin: 28 de marzo del 2018
Descripción: El usuario accede al módulo Temáticas donde puede entrar a cada una de las temáticas que se registraron en el sistema, haciendo clic sobre alguna, se ejecuta la vista mostrarplantilla pasándole por parámetro el identificador de la temática que se desea mostrar, renderizando sus datos en una página permitiendo al usuario observar el análisis de los resultados de la búsqueda.	

Tabla 36. Tarea de Ingeniería - Modificar clasificación.

Tarea de Ingeniería	
Nombre: Modificar clasificación.	
HU: Gestionar temáticas.	Puntos Estimados: 0.6
Fecha de inicio: 29 de marzo del 2018	Fecha fin: 2 de abril del 2018
Descripción: Dentro de cada temática accedida, el usuario puede modificar el estado de la clasificación de los resultados de la búsqueda, seleccionando otra clasificación dentro de las posibles a elegir y haciendo clic en reclasificar se envía como parámetro la nueva clasificación a la vista reclasificar, la cual implementa la lógica difusa cambiando el peso de los selectores para las páginas que se vinculan a ellos y realizando un nuevo análisis de los resultados, el cual se muestra al usuario recargando la interfaz correspondiente a la temática en la que está trabajando.	

Tabla 37. Tarea de Ingeniería - Eliminar temáticas.

Tarea de Ingeniería	
Nombre: Eliminar temáticas.	



HU: Gestionar temáticas.	Puntos Estimados: 0.2
Fecha de inicio: 3 de abril del 2018	Fecha fin: 3 de abril del 2018
Descripción: El usuario puede eliminar una temática de su listado, haciendo clic en Eliminar Temática, lo cual ejecuta la vista eliminar temática pasándole por parámetro el id de la temática a eliminar. Esta acción debe confirmarse debido a que al eliminar una temática se eliminarían todas las URLs, selectores y búsquedas asociadas a ella, siendo irreversible esta acción.	

- HU-Configurar automáticamente el monitoreo.

Tabla 38. Tarea de Ingeniería - Configurar automáticamente el monitoreo.

Tarea de Ingeniería	
Nombre: Configurar automáticamente el monitoreo.	
HU: Configurar automáticamente el monitoreo.	Puntos Estimados: 1
Fecha de inicio: 4 de abril del 2018	Fecha fin: 10 de abril del 2018
Descripción: El usuario accede al módulo URLs donde haciendo clic en monitoreo automático el sistema debe permitir configurar el monitoreo de las URLs de manera automática por medio de tareas programadas que el sistema ejecute en un tiempo determinado por el usuario, esta tarea inicia el monitoreo siguiendo los mismos pasos que de forma manual.	

Pruebas unitarias

En esta iteración se realizaron 18 pruebas unitarias al sistema. Los resultados arrojados pueden observarse a continuación:



```

Creating test database for alias 'default'...
System check identified no issues (0 silenced).
.....
-----
Ran 18 tests in 0.344s
OK
Destroying test database for alias 'default'...
    
```

Fig. 15 Resultados de las pruebas unitarias de la tercera iteración

3.6 Iteración 4

Durante esta iteración se abordaron las HU Gestión de selectores, Gestión de URL y Mostrar el estado de ejecución de la búsqueda y sus trazas de arranque y parada.

- HU-Gestionar selectores.

Tabla 39. Tarea de Ingeniería - Añadir selectores.

Tarea de Ingeniería	
Nombre: Añadir selectores.	
HU: Gestionar selectores.	Puntos Estimados: 0.2
Fecha de inicio: 11 de abril del 2018	Fecha fin: 11 de abril del 2018
Descripción: El usuario accede al módulo Selectores donde haciendo clic en agregar selector se muestra un formulario donde se entran los campos que identificarán al selector, nombre, peso y temática a la que pertenece, al aceptar, se ejecuta la vista agregar_selector la cual recibe como parámetros estos datos y agrega el selector a la base de datos.	

Tabla 40. Tarea de Ingeniería - Modificar selectores.

Tarea de Ingeniería	
Nombre: Modificar selectores.	
HU: Gestionar selectores.	Puntos Estimados: 0.2



Fecha de inicio: 12 de abril del 2018	Fecha fin: 12 de abril del 2018
Descripción: El usuario accediendo al módulo Selectores puede seleccionar la opción modificar selector de los selectores listados, donde haciendo clic sobre ella muestra un formulario con los campos de entrada para modificar el selector, al aceptar, se ejecuta la vista update_select la cual recibe como parámetro el id del selector seleccionado y sus nuevos datos haciendo su modificación en la base de datos.	

Tabla 41. Tarea de Ingeniería - Mostrar selectores.

Tarea de Ingeniería	
Nombre: Mostrar selectores.	
HU: Gestionar selectores.	Puntos Estimados: 0.2
Fecha de inicio: 13 de abril del 2018	Fecha fin: 13 de abril del 2018
Descripción: El sistema renderiza a través de la vista selectores una página donde muestra en una tabla los selectores listados.	

Tabla 42. Tarea de Ingeniería - Listar selectores.

Tarea de Ingeniería	
Nombre: Listar selectores.	
HU: Gestionar selectores.	Puntos Estimados: 0.2
Fecha de inicio: 16 de abril del 2018	Fecha fin: 16 de abril del 2018
Descripción: Al acceder al módulo Selectores, el sistema ejecuta la vista selectores listando todos los que se encuentran en la base de datos y los agrupa por la temática a la que pertenecen.	

Tabla 43. Tarea de Ingeniería - Eliminar selectores.

Tarea de Ingeniería	
Nombre: Eliminar selectores.	
HU: Gestionar selectores.	Puntos Estimados: 0.2
Fecha de inicio: 17 de abril del 2018	Fecha fin: 17 de abril del 2018
Descripción: El usuario accediendo al módulo Selectores puede seleccionar la opción eliminar selector de los selectores listados, donde haciendo clic sobre ella se ejecuta la vista message la cual recibe como parámetro el id del selector seleccionado y muestra un modelo con un mensaje de seguridad para la eliminación del mismo, si el mensaje es aceptado el sistema ejecuta la vista elim_select y lo elimina de la base de datos.	



➤ HU-Gestionar URLs.

Tabla 44. Tarea de Ingeniería - Añadir URL.

Tarea de Ingeniería	
Nombre: Añadir URL.	
HU: Gestionar URLs.	Puntos Estimados: 0.4
Fecha de inicio: 18 de abril del 2018	Fecha fin: 19 de abril del 2018
Descripción: Al acceder al módulo URLs, el usuario puede agregar una URL de forma manual, al presionar agregar URL se muestra un formulario donde se pueden entrar los campos para la nueva URL, cuando se aceptan estas entradas se ejecuta la vista add URL, la cual recibe desde el formulario como parámetros los datos para conformar la URL agregándola a la base de datos.	

Tabla 45. Tarea de Ingeniería - Modificar URL.

Tarea de Ingeniería	
Nombre: Modificar URL.	
HU: Gestionar URLs.	Puntos Estimados: 0.4
Fecha de inicio: 20 de abril del 2018	Fecha fin: 23 de abril del 2018
Descripción: El usuario accediendo al módulo URLs puede seleccionar la opción modificar URL de las URLs listadas, donde haciendo clic sobre ella se ejecuta la vista mod_ URL la cual recibe como parámetro el id de la URL a modificar y muestra un formulario con los campos de entrada para modificarla, al aceptar se ejecuta la vista update_ URL, la cual recibe como parámetro las entradas de los campos a modificar y realiza la modificación en la base de datos.	

Tabla 46. Tarea de Ingeniería - Mostrar URL.

Tarea de Ingeniería	
Nombre: Mostrar URL.	
HU: Gestionar URLs.	Puntos Estimados: 0.2
Fecha de inicio: 24 de abril del 2018	Fecha fin: 24 de abril del 2018
Descripción: El sistema renderiza a través de la vista URLs una página donde muestra en una tabla las URL listadas.	



Tabla 47. Tarea de Ingeniería - Listar URL.

Tarea de Ingeniería	
Nombre: Listar URL.	
HU: Gestionar URLs.	Puntos Estimados: 0.2
Fecha de inicio: 25 de abril del 2018	Fecha fin: 25 de abril del 2018
Descripción: Al acceder al módulo URLs, el sistema ejecuta la vista URL listando todos las URLs que se encuentran en la base de datos, agrupándolas por la temática a la que pertenecen haciendo una petición a la base de datos de todas las URLs asociadas a cada temática.	

Tabla 48. Tarea de Ingeniería - Eliminar URL.

Tarea de Ingeniería	
Nombre: Eliminar URL.	
HU: Gestionar URLs.	Puntos Estimados: 0.2
Fecha de inicio: 26 de abril del 2018	Fecha fin: 26 de abril del 2018
Descripción: El usuario accediendo al módulo URLs puede seleccionar la opción eliminar URL de las URLs listadas, donde haciendo clic sobre alguna de ella se ejecuta la vista mensaje, la cual recibe como parámetro el id de la URL seleccionada y muestra un modelo con un mensaje de seguridad para la eliminación de la misma, si el mensaje es aceptado, el sistema ejecuta la vista eliminar_ URL y la elimina de la base de datos.	

- HU- Listar el registro de la búsqueda y el número de URLs analizadas.

Tabla 49. Tarea de Ingeniería - *Listar el registro de la búsqueda y el número de URLs analizadas.*

Tarea de Ingeniería	
Nombre: Listar el registro de la búsqueda y el número de URLs analizadas.	
HU: Listar el registro de la búsqueda y el número de URLs analizadas.	Puntos Estimados: 0.6
Fecha de inicio: 27 de abril del 2018	Fecha fin: 1 de mayo del 2018
Descripción: Al iniciar un monitoreo de forma manual, el sistema ejecuta la vista show_bar mostrando una barra de estado, esta posibilita al usuario observar el porcentaje por donde va la búsqueda en la red.	

Pruebas unitarias



En esta iteración se realizaron 32 pruebas unitarias al sistema. Los resultados pueden observarse en la siguiente figura:

```

Creating test database for alias 'default'...
System check identified no issues (0 silenced).
.....
-----
Ran 32 tests in 0.604s

OK
Destroying test database for alias 'default'...
    
```

Fig. 16 Resultados de las pruebas unitarias de la cuarta iteración

3.7 Iteración 5

En esta iteración se llevó a cabo la implementación de las HU Generar documento PDF con resultados seleccionados, Gestionar envío de informes por correo y Gestionar envío de informes por correo.

- HU-Generar documento PDF con resultados seleccionados.

Tabla 50. Tarea de Ingeniería - Generar documento PDF con los resultados seleccionados

Tarea de Ingeniería	
Nombre: Generar documento PDF con resultados seleccionados.	
HU: Generar documento PDF con resultados seleccionados.	Puntos Estimados: 0.6
Fecha de inicio: 2 de mayo del 2018	Fecha fin: 4 de mayo del 2018
Descripción: El sistema debe ser capaz de generar un documento PDF donde se puedan recoger los resultados de las búsquedas más importantes para el usuario y así tener un resumen de contenidos que se puedan adjuntar a un correo electrónico o se puedan exportar.	

- HU-Authenticar usuario.



Tabla 51. Tarea de Ingeniería - Autenticarse usuario.

Tarea de Ingeniería	
Nombre: Autenticar usuario.	
HU: Autenticarse usuario.	Puntos Estimados: 0.4
Fecha de inicio: 7 de mayo del 2018	Fecha fin: 8 de mayo del 2018
<p>Descripción: El usuario accede a la página de inicio del sistema donde se le pedirá entrar un usuario y contraseña a través de un formulario, al aceptar, se ejecuta la vista index_log la cual verifica la autenticidad de las credenciales entradas, lanzando un mensaje de error si no son aceptadas. No se permitirá la entrada a la página principal del sistema hasta que no se hayan aceptado los datos del usuario.</p>	

- HU- Configurar envío de informes por correo.

Tabla 52. Tarea de Ingeniería - Gestionar envío de informes por correo.

Tarea de Ingeniería	
Nombre: Configurar envío de informes por correo.	
HU: Configurar envío de informes por correo.	Puntos Estimados: 1
Fecha de inicio: 9 de mayo del 2018	Fecha fin: 15 de mayo del 2018
<p>Descripción: El usuario al acceder al módulo de temas y entrando a un tema en específico, puede enviar un correo electrónico con los datos del análisis de los resultados de la búsqueda, haciendo clic en enviar correo, el sistema ejecuta la vista send_email recibiendo como parámetros el id del tema del que se quieren mostrar los resultados, muestra una página donde se completan los campos de dirección de correo a enviar asunto y cuerpo, este último dato se rellena de manera automática con los datos asociados al tema.</p>	

- HU- Configurar envío de forma automática de los resultados.

Tabla 53. Tarea de Ingeniería - Envío de forma automática de los resultados.

Tarea de Ingeniería	
Nombre: Configurar envío de forma automática de los resultados.	
HU: Configurar envío de forma automática de los resultados.	Puntos Estimados: 1
Fecha de inicio: 16 de mayo del 2018	Fecha fin: 22 de mayo del 2018



Descripción: El usuario al acceder al módulo de temas y entrando a un tema en específico, puede configurar el envío de los correos electrónicos de manera automática al Presionar envío automático, el sistema ejecuta la vista `send_email_auto` la cual recibe como parámetro el id del tema del que se quieren mostrar los resultados y muestra un formulario donde se pueden entrar los datos de fecha, hora, destinatario y asunto para los correos que enviará de manera automática. El sistema ejecuta el envío de los correos de la misma manera que de forma manual, al comprobar que la fecha y hora configuradas son las del sistema, llenando el cuerpo del correo con los datos actuales que posea el tema para el cual se configuró.

Pruebas unitarias

En esta iteración se realizaron 38 pruebas unitarias al sistema. Los resultados pueden observarse a continuación:

```

Creating test database for alias 'default'...
System check identified no issues (0 silenced).
.....
-----
Ran 38 tests in 0.692s

OK
Destroying test database for alias 'default'...
    
```

Fig. 17 Resultados de las pruebas unitarias de la quinta iteración

En la siguiente gráfica se muestran los resultados de los casos de prueba de aceptación en cada iteración:

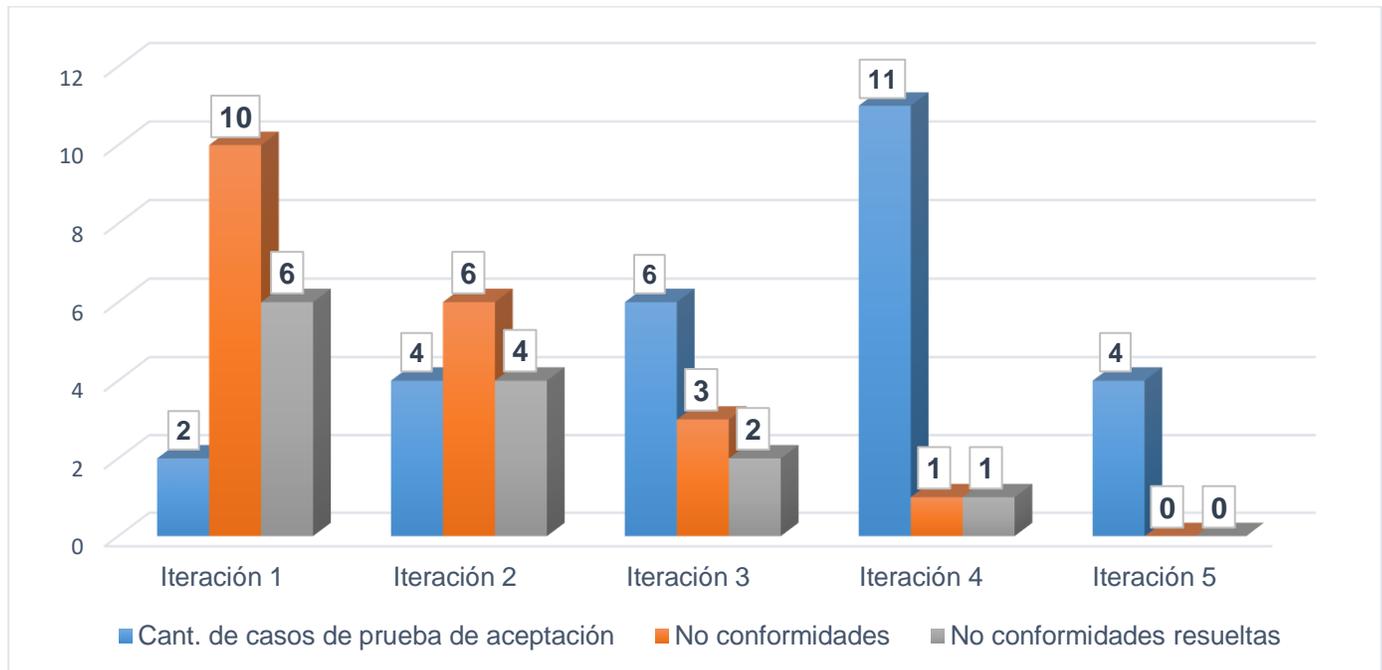


Fig. 18 Resultados de los Casos de Pruebas de Aceptación

3.8 Comprobación del aprendizaje del algoritmo

Para comprobar la técnica de inteligencia artificial utilizada en esta aplicación se crearon dos temas principales acordes a temáticas de interés para el centro, **GRHS** y **Call Center**, monitoreándose sitios web de productos similares a los del centro que brindan canales RSS para consultar los nuevos artículos que publican, como por ejemplo <https://www.fiixsoftware.com/feed/> y <https://www.nextiva.com/blog/feed>. Las URLs de estos sitios web se asociaron a estos temas, además se crearon 9 selectores que se aplicarían a la información arrojada por dichas búsquedas 5 para la temática **Call Center**: *recording, call, dialer, conferencing, service* y 4 para **GRHS**: *hardware, software, management y asset*. De ahí se obtuvieron con el monitoreo un conjunto de resultados que fueron analizados con estos selectores, aplicando una función que retornó la relevancia que posee el bloque de información.

Esta función realizó una media de los pesos de los selectores que aparecen en los bloques de información analizada por temática, además se analizó la cantidad de apariciones que tuvo cada selector dentro de todos los bloques de información analizados para cada temática. El resultado para diferentes bloques de información, donde aparecieron selectores asociados a la temática fue el siguiente:

Temática: *Call Center*

URL: <https://www.nextiva.com/blog/feed>



Tabla 54. Primer resultado para la búsqueda de la temática *Call Center*

Bloque de información	Selectores Encontrados en el bloque.	Clasificación del sistema para el bloque.	Reclasificación del usuario
1	Service	Baja	Baja
2	service, conferencing	Media	Alta
3	service,	Media	Media

Al realizarse la reclasificación por el usuario atendiendo al interés que tuvo en cada bloque de información que se le brindó, el sistema hizo una base de conocimientos poblada con los selectores que se analizaron en ese bloque de información y sus pesos correspondientes, la dirección URL de donde se obtuvo la información y un nuevo peso asociado a cada selector recalculado según la clasificación reasignada por el usuario.

El sistema aprendió de esta decisión del usuario, que los selectores que se utilizaron para clasificar la información obtenida desde esa URL presentan un nuevo peso e importancia en la búsqueda y recuperación de la información.

Al realizarse una nueva búsqueda con nuevos resultados, en ese sitio web, la clasificación de la información obtenida se realizó atendiendo a las nuevas entradas para los pesos. Además de que el sistema mostró los bloques de información con mayor presencia de selectores de mayor peso como primeros resultados quedando de esta manera:

Temática: *Call Center*

URL: <https://www.nextiva.com/blog/feed>

Tabla 55. Segundo resultado para la búsqueda de la temática *Call Center*

Bloque de información	Selectores Encontrados en el bloque.	Clasificación del sistema para el bloque.	Reclasificación del usuario
1	<i>conferencing</i>	Media	-
2	<i>conferencing</i>	Media	-
3	<i>service</i>	Media	Alta



El resultado que se puede observar es que para los bloques de información dentro de la temática **Call Center**, los selectores *service* y *conferencing*, presentan un mayor grado de aparición, además de que la clasificación de esa información donde aparecen, es de relevancia para el usuario y dentro de la temática **GRHS** los que mayor presencia y relevancia poseen para el usuario son *asset* y *management*.

3.9 Conclusiones del capítulo

El diseño arquitectónico es la base para el óptimo desarrollo de un sistema, por este motivo la presente investigación respalda su diseño sobre patrones bien establecidos. La metodología propuesta presenta sus métodos propios para la confección de las tareas de diseño, usando para este fin las tarjetas CRC con el objetivo de especificar las responsabilidades de cada clase presente en la lógica del negocio. Además, las tareas de implementación ejecutadas sirvieron como soporte organizativo en el desarrollo de la codificación del software propuesto y posibilitaron distribuir de forma organizada la programación entre los miembros del equipo de desarrollo. Una de las ventajas que provee la metodología XP al efectuar las pruebas, es la posibilidad de que una vez expuesto el avance de la aplicación al cliente, este puede identificar las inconformidades existentes, las cuales el equipo de desarrollo corrige de inmediato.



CONCLUSIONES

Con la culminación del trabajo de diploma Sistema Informático para la Vigilancia Tecnológica del Centro de Telemática de la Universidad de la Ciencias Informáticas, se dio cumplimiento al objetivo propuesto en la investigación, obteniendo una solución que responde a las necesidades identificadas en el proceso de Mercadotecnia del Centro de Telemática.

El análisis del ambiente nacional e internacional, permitió reconocer la no existencia de una solución informática que diese soporte a las necesidades planteadas por el cliente. A partir de ello, se propuso una solución que tuvo un modelo guía mediante la implementación de los artefactos correspondientes a los flujos de trabajo propuestos por la metodología XP.

La construcción de la solución propuesta empleando *web scraping* y lógica difusa viabiliza el proceso de Mercadotecnia al permitir la optimización de los tiempos de las búsquedas y garantizar la recopilación de información actualizada y precisa, lo que implica una notable mejora para el desarrollo de la vigilancia tecnológica en el centro TLM y, por consiguiente, se elevan la calidad y competitividad de su cartera de productos y servicios.

Las pruebas realizadas a la solución permitieron identificar los errores existentes en la misma, arrojando no conformidades que fueron corregidas en su totalidad, dándole validez, estabilidad y confiabilidad al sistema implementado.



RECOMENDACIONES

Se recomienda continuar con la evolución de la solución desarrollada, agregando nuevas funcionalidades, como son:

- Ampliar la recuperación de información, con la implementación de búsquedas en documentos con formatos DOC y PDF, presentes en la web.
- Incluir la extracción de metadatos de archivos de bases de datos *Open Access*, a través del protocolo OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*).
- Extender la configuración del monitoreo periódico, permitiendo la selección de varias URLs por temática.

REFERENCIAS BIBLIOGRÁFICAS

- Robles Garrote, P. y Rojas, M. D. C. 2015.** *La validación por juicio de expertos: dos investigaciones cualitativas en Lingüística aplicada.* 2015.
- A Kenneth Reitz Project. 2013.** Requests. [En línea] 2013. [Citado el: 12 de 03 de 2018.] <http://es.python-requests.org/es/latest/>.
- AIMPLAS. 2017.** sofVT. [En línea] 2017. [Citado el: 05 de 01 de 2017.] <http://www.softvt.com/que-es-softvt.php>.
- 1998-2018.** ALEGSA.com.ar. *ALEGSA.com.ar.* [En línea] 1998-2018. [Citado el: 01 de 04 de 2018.] http://www.alegsa.com.ar/Dic/validacion_y_prueba_de_sistemas.php.
- 2018.** Asier Marqués. *Asier Marqués.* [En línea] 2018. <http://asiermarques.com/2013/conceptos-sobre-apis-rest/i..>
- Beck, Kent. 2000.** *Extreme Programming Explained. Embrace Change.* 2000.
- Becker, Kent. 1999.** *Extreme Programming Explained.* 1999.
- Benítez, Raúl, Escudero, Gerardo y Kanaan, Samir. 2013.** *Inteligencia artificial avanzada.* Barcelona : UOC, 2013. 978-84-9064-321-1.
- Berzal, Fernando, Martín-Bautista, María J. y Blanco, Ignacio J. 2002.** *La lógica difusa en el proceso de recuperación de Información.* 2002.
- Bordogna, G. , Carrara , P. y Pasi, G. 1995.** *Fuzzy Approaches to Extend Boolean Information Retrieval. En Bosc. P y Kacprzyk, J. Fuzziness in Database Management Systems.* Germany : s.n., 1995.
- Bustamante, Dayana y Rodríguez, Jean C. Marzo del 2014.** *Metodología Actual. Metodología XP.* Barinas : s.n., Marzo del 2014.
- CDE. Hontza. [En línea] [Citado el: 12 de 01 de 2017.]** <http://www.hontza.es/es>.
- Cockburn, A. 2000.** *Selecting a project's methodology. IEEE Software.* 2000.
- Craig, Larman. 2001.** *UML y Patrones. 2. .* 2001.
- Definición.Es. 2008-2018.** Definición.Es. [En línea] 2008-2018. [Citado el: 14 de 03 de 2018.] <https://definicion.de/estandar/>.
- .** **2008-2018.** Definición.Es. [En línea] 2008-2018. [Citado el: 16 de 03 de 2018.] <https://definicion.de/codificacion/>.
- Demir, Begum y Brusone, Lorenzo. 2015.** *A novel active learning method in relevance feedback for content-based remote sensing image retrieval.* EE.UU : IEEE, 2015. 0196-2892.
- Dirección de Comunicación Institucional, UCI. 2016.** *Plan de Comunicación Informática 2016.* La Habana : s.n., 2016.

- 2005-2018.** django. *django*. [En línea] 2005-2018. [Citado el: 26 de 02 de 2018.] <https://docs.djangoproject.com/es/2.0/faq/general/>.
- 2018.** Django REST framework. *Django REST framework*. [En línea] 2018. <http://www.django-rest-framework.org/>.
- Django Software Foundation. 2005-2018.** django. [En línea] 2005-2018. [Citado el: 16 de 11 de 2017.] <https://www.djangoproject.com>.
- Elección entre una metodología ágil y tradicional basado en técnicas de soft computing.* **Crésopo Boaventura , José, y otros. 2016.** 2016, Vol. 10. 2227-1899.
- 2018.** Exploiting web scraping in a collaborative filteringbased approach to web advertising. Artificial Intelligence Research . [En línea] 2018. [Citado el: 04 de 12 de 2016.] [http://www.sciedu.ca/journal/index.php/air/..](http://www.sciedu.ca/journal/index.php/air/)
- 2009-2017.** Gestión-Calidad.com. *Gestión-Calidad.com*. [En línea] 2009-2017. [Citado el: 4 de 11 de 2017.] <http://gestion-calidad.com/vigilancia-tecnologica>.
- Gonçalves Torres, Luis Manuel. 2015.** *HERRAMIENTA DE CONSUMO Y REUTILIZACIÓN DE CONTENIDO WEB*. La Habana : s.n., 2015.
- 2013.** ingenio DS. *ingenio DS*. [En línea] 2013. [Citado el: 24 de 02 de 2017.] <https://ingeniods.wordpress.com/2013/09/16/patrones-arquitectonicos/>.
- Introducción al razonamiento aproximado: lógica difusa.* **De Negri, Carlos Eduardo y De Vito, Eduardo Luis. 2006.** 4, 2006. 126-136.
- JetBrains s.r.o. . 2000 - 2018.** PyCharm. [En línea] 2000 - 2018. [Citado el: 15 de 03 de 2018.] [https://www.jetbrains.com/pycharm/..](https://www.jetbrains.com/pycharm/)
- Joskowicz, Ing. José. 2008.** *Reglas y prácticas en Extreme Programming*. 2008.
- 2012.** Jummp. *Jummp*. [En línea] 2012. [Citado el: 26 de 02 de 2018.] <https://jummp.wordpress.com/2012/01/10/desarrollo-de-software-tarjetas-crc/>.
- La vigilancia tecnológica en la gestión de proyectos de I+D+i: recursos y herramientas.* **Muñoz Durán, Javier , Marín Martínez, María y Vallejo Triano, José. 2006.** 6, 2006, Vol. 15.
- Mabrouk, Dooa, Rady, Sherine y Bard, Nagwa. 2017.** *A survey information retrieval systems´ modeling using term dependencies and term weighting*. EE.UU : IEEE, 2017. 978-1-5386-0821-0.
- Martinez Guerrero, Rafael. 2009-2013.** PostgreSQL-es. [En línea] 2009-2013. www.postgresql.org.es.
- 2016.** *METODOLOGIA ÁGIL DE DESARROLLO DE SOFTWARE PROGRAMACION*. 2016.
- Mozilla y colaboradores individuales. 2005-2018.** MDN web docs. [En línea] 2005-2018. <https://developer.mozilla.org/es/docs/Web/CSS>.

- . **2005-2018.** MDN web docs. [En línea] 2005-2018.
https://developer.mozilla.org/es/docs/Web/JavaScript/Acerca_de_JavaScript.
- Nicholls, Tom y Bright, Jonathan. 2018.** *Understanding news story chains using retrieval and network clustering techniques.* EEUU : arXiv.org, 2018. 1801.07988v1.
- Octopus Data Inc. 2017.** Octoparse. [En línea] 2017. <http://www.octoparse.com/>.
On the Delusiveness of Adapting a Common Space for Modeling IR Objects: Are Queries Documents?
- Bollman , P. y Raghavan, V.V. 1993.** 10, 1993, Vol. 44.
- Otto, Mark. B.** [En línea] [Citado el: 23 de 01 de 2018.] <https://getbootstrap.com/>.
- OVTT.** Observatorio Virtual de Transferencia de Tecnología . [En línea] [Citado el: 28 de 01 de 2018.]
<http://www.ovtt.org/vigilancia-tecnologica>.
- Pilgrim, Mark. Agosto 2010.** *Html5 Up and running: Dive in to the future of the web developing.* . Agosto 2010.
- 2018.** PMOinformatica.com. *PMOinformatica.com.* [En línea] 2018. [Citado el: 12 de 02 de 2017.]
<http://www.pmoinformatica.com/2015/05/requerimientos-no-funcionales-ejemplos.html>.
- 2018.** PMOinformatica.com. *PMOinformatica.com.* [En línea] 2018. [Citado el: 13 de 02 de 2017.]
<http://www.pmoinformatica.com/2017/02/requerimientos-funcionales-ejemplos.html>.
- Pressman, R.S. 2005.** *Software Engineering: A Practitioner's Approach.* . 2005.
- 2018.** Programarfacil.com. *Programarfacil.com.* [En línea] 2018. [Citado el: 20 de 02 de 2018.]
<https://programarfacil.com/blog/programacion/los-patrones-de-diseno/>.
- 2001-2018.** python. [En línea] 2001-2018. [Citado el: 01 de 03 de 2017.] <https://www.python.org/>.
- Python Software Foundation (US) . 2001 - 2018.** python. [En línea] 2001 - 2018. [Citado el: 12 de 02 de 2017.] <https://www.python.org/>.
- Python Software Foundation (US). 20001-2018.** phyton. [En línea] 20001-2018. [Citado el: 20 de 03 de 2018.] <https://www.python.org/dev/peps/pep-0008/>.
- Python Software Foundation. 1990-2018.** python. [En línea] 1990-2018. [Citado el: 12 de 04 de 2018.]
<https://docs.python.org/2/library/unittest.html?highlight=uni#module-unittest>.
- Richardson, Leonard. 1998-2018.** Beautiful Soup. [En línea] 1998-2018. [Citado el: 15 de 03 de 2018.]
<https://www.crummy.com/software/BeautifulSoup/>.
- S.P.Ruba, Rani, Ramesh, B y Anusha, M. 2015.** *Evaluation of stemming techniques for text classification.*
s.l. : International Journal of Computer Science and Mobile Computing, 2015. 2320-088X.
- Scraping, Web. 2014.** *What is import·io from the user's point of view?* [En línea] 2014. [Citado el: 14 de 11 de 2017.] <http://scraping.pro/what-is-import-io/>.

- Shaw, M. y Garlan, D. 1995.** *“Formulations and Formalisms in Software Architecture”*. s.l. : Springer-Verlag, 1995. Volume 1000—Lecture Notes in Computer Science.
- SOME ASPECTS REGARDING ON TECHNOLOGICAL VIGILA. Savescu, Dan. 2014.* 2014.
- Tecnología IALE. 2016.** Vigiale. [En línea] 2016. [Citado el: 21 de 11 de 2017.] <https://www.vigiale.com/>.
- The jQuery Foundation. 2018.** jQuery. [En línea] 2018. <https://jquery.com/>.
- The Use of Web-scraping Software in Searching for Grey Literature. Haddaway, Neal R. 2015.* 3, 2015, Vol. 11.
- 2017.** UniMOOC. *UniMOOC*. [En línea] 2017. <https://unimooc.com/>.
- Web Scraping. 2002 - 2014.** Helium Scraper Review. [En línea] 2002 - 2014. [Citado el: 15 de 11 de 2017.] <http://scraping.pro/helium-scraper-review/>.
- . **2002 - 2014.** OutWit Hub Review. [En línea] 2002 - 2014. [Citado el: 16 de 11 de 2017.] <http://scraping.pro/outwit-hub-review/>.
- . **2002 - 2014.** Software for Web Scraping. [En línea] 2002 - 2014. [Citado el: 05 de 11 de 2017.] <http://scraping.pro/software-for-web-scraping/>.
- . **2002 - 2014.** Web Harvy Data Extractor. [En línea] 2002 - 2014. [Citado el: 17 de 11 de 2017.] <http://scraping.pro/sysnucleus-webharvy/>.
- . **2002 - 2014.** What is import.io from the user`s point of view. [En línea] 2002 - 2014. [Citado el: 01 de 02 de 2017.] <http://scraping.pro/what-is-import-io/>.
- WebExtractor System. 2018.** Web Data Extractor. Automatically extract specific information from web pages. [En línea] 2018. [Citado el: 15 de 06 de 2017.] <http://www.webextractor.com/help.htm>.
- Wusteman, Judith. 2004.** *RSS: the latest feed*. s.l. : Emerald Group Publishing Limited, 2004.
- Yao, Y.Y. 1995.** *Measuring Retrieval Effectiveness Based on User Preference of Documents*. *Journal of the American Society for Information Science*. 1995.

BIBLIOGRAFÍA

- Robles Garrote, P. y Rojas, M. D. C. 2015.** *La validación por juicio de expertos: dos investigaciones cualitativas en Lingüística aplicada.* 2015.
- A Kenneth Reitz Project. 2013.** Requests. [En línea] 2013. [Citado el: 12 de 03 de 2018.] <http://es.python-requests.org/es/latest/>.
- AIMPLAS. 2017.** sofVT. [En línea] 2017. [Citado el: 05 de 01 de 2017.] <http://www.softvt.com/que-es-softvt.php>.
- 1998-2018.** ALEGSA.com.ar. *ALEGSA.com.ar.* [En línea] 1998-2018. [Citado el: 01 de 04 de 2018.] http://www.alegsa.com.ar/Dic/validacion_y_prueba_de_sistemas.php.
- 2018.** Asier Marqués. *Asier Marqués.* [En línea] 2018. <http://asiermarques.com/2013/conceptos-sobre-apis-rest/i..>
- Beck, Kent. 2000.** *Extreme Programming Explained. Embrace Change.* 2000.
- Becker, Kent. 1999.** *Extreme Programming Explained.* 1999.
- Benítez, Raúl, Escudero, Gerardo y Kanaan, Samir. 2013.** *Inteligencia artificial avanzada.* Barcelona : UOC, 2013. 978-84-9064-321-1.
- Berzal, Fernando, Martín-Bautista, María J. y Blanco, Ignacio J. 2002.** *La lógica difusa en el proceso de recuperación de Información.* 2002.
- Bordogna, G. , Carrara , P. y Pasi, G. 1995.** *Fuzzy Approaches to Extend Boolean Information Retrieval. En Bosc. P y Kacprzyk, J. Fuzziness in Database Management Systems.* Germany : s.n., 1995.
- Bustamante, Dayana y Rodríguez, Jean C. Marzo del 2014.** *Metodología Actual. Metodología XP.* Barinas : s.n., Marzo del 2014.
- CDE. Hontza.** [En línea] [Citado el: 12 de 01 de 2017.] <http://www.hontza.es/es>.
- Cockburn, A. 2000.** *Selecting a project's methodology. IEEE Software.* 2000.
- Craig, Larman. 2001.** *UML y Patrones. 2. .* 2001.
- Definición.Es. 2008-2018.** Definición.Es. [En línea] 2008-2018. [Citado el: 14 de 03 de 2018.] <https://definicion.de/estandar/>.
- . 2008-2018.** Definición.Es. [En línea] 2008-2018. [Citado el: 16 de 03 de 2018.] <https://definicion.de/codificacion/>.
- Demir, Begum y Brusone, Lorenzo. 2015.** *A novel active learning method in relevance feedback for content-based remote sensing image retrieval.* EE.UU : IEEE, 2015. 0196-2892.
- Dirección de Comunicación Institucional, UCI. 2016.** *Plan de Comunicación Informática 2016.* La Habana : s.n., 2016.

- 2005-2018.** django. *django*. [En línea] 2005-2018. [Citado el: 26 de 02 de 2018.]
<https://docs.djangoproject.com/es/2.0/faq/general/>.
- 2018.** Django REST framework. *Django REST framework*. [En línea] 2018. <http://www.django-rest-framework.org/>.
- Django Software Foundation. 2005-2018.** django. [En línea] 2005-2018. [Citado el: 16 de 11 de 2017.]
<https://www.djangoproject.com>.
- Elección entre una metodología ágil y tradicional basado en técnicas de soft computing.* **Crêspo Boaventura , José, y otros. 2016.** 2016, Vol. 10. 2227-1899.
- 2018.** Exploiting web scraping in a collaborative filteringbased approach to web advertising. Artificial Intelligence Research . [En línea] 2018. [Citado el: 04 de 12 de 2016.]
[http://www.sciedu.ca/journal/index.php/air/..](http://www.sciedu.ca/journal/index.php/air/)
- 2009-2017.** Gestión-Calidad.com. *Gestión-Calidad.com*. [En línea] 2009-2017. [Citado el: 4 de 11 de 2017.] <http://gestion-calidad.com/vigilancia-tecnologica>.
- Gonçalves Torres, Luis Manuel. 2015.** *HERRAMIENTA DE CONSUMO Y REUTILIZACIÓN DE CONTENIDO WEB*. La Habana : s.n., 2015.
- 2013.** ingenio DS. *ingenio DS*. [En línea] 2013. [Citado el: 24 de 02 de 2017.]
<https://ingeniods.wordpress.com/2013/09/16/patrones-arquitectonicos/>.
- Introducción al razonamiento aproximado: lógica difusa.* **De Negri, Carlos Eduardo y De Vito, Eduardo Luis. 2006.** 4, 2006. 126-136.
- JetBrains s.r.o. . 2000 - 2018.** PyCharm. [En línea] 2000 - 2018. [Citado el: 15 de 03 de 2018.]
[https://www.jetbrains.com/pycharm/..](https://www.jetbrains.com/pycharm/)
- Joskowicz, Ing. José. 2008.** *Reglas y prácticas en Extreme Programing*. 2008.
- 2012.** Jummp. *Jummp*. [En línea] 2012. [Citado el: 26 de 02 de 2018.]
<https://jummp.wordpress.com/2012/01/10/desarrollo-de-software-tarjetas-crc/>.
- La vigilancia tecnológica en la gestión de proyectos de I+D+i: recursos y herramientas.* **Muñoz Durán, Javier , Marín Martínez, María y Vallejo Triano, José. 2006.** 6, 2006, Vol. 15.
- Mabrouk, Dooa, Rady, Sherine y Bard, Nagwa. 2017.** *A survey information retrieval systems´ modeling using term dependencies and term weighting*. EE.UU : IEEE, 2017. 978-1-5386-0821-0.
- Martinez Guerrero, Rafael. 2009-2013.** PostgreSQL-es. [En línea] 2009-2013. www.postgresql.org.es.
- 2016.** *METODOLOGIA ÁGIL DE DESARROLLO DE SOFTWARE PROGRAMACION*. 2016.
- Mozilla y colaboradores individuales. 2005-2018.** MDN web docs. [En línea] 2005-2018.
<https://developer.mozilla.org/es/docs/Web/CSS>.

- . 2005-2018. MDN web docs. [En línea] 2005-2018.
https://developer.mozilla.org/es/docs/Web/JavaScript/Acerca_de_JavaScript.
- Nicholls, Tom y Bright, Jonathan. 2018.** *Understanding news story chains using retrieval and network clustering techniques*. EEUU : arXiv.org, 2018. 1801.07988v1.
- Octopus Data Inc. 2017.** Octoparse. [En línea] 2017. <http://www.octoparse.com/>.
On the Delusiveness of Adapting a Common Space for Modeling IR Objects: Are Queries Documents?
- Bollman , P. y Raghavan, V.V. 1993.** 10, 1993, Vol. 44.
- Otto, Mark. B.** [En línea] [Citado el: 23 de 01 de 2018.] <https://getbootstrap.com/>.
- OVTT.** Observatorio Virtual de Transferencia de Tecnología . [En línea] [Citado el: 28 de 01 de 2018.]
<http://www.ovtt.org/vigilancia-tecnologica>.
- Pilgrim, Mark. Agosto 2010.** *Html5 Up and running: Dive in to the future of the web developing*. . Agosto 2010.
- 2018.** PMOinformatica.com. *PMOinformatica.com*. [En línea] 2018. [Citado el: 12 de 02 de 2017.]
<http://www.pmoinformatica.com/2015/05/requerimientos-no-funcionales-ejemplos.html>.
- 2018.** PMOinformatica.com. *PMOinformatica.com*. [En línea] 2018. [Citado el: 13 de 02 de 2017.]
<http://www.pmoinformatica.com/2017/02/requerimientos-funcionales-ejemplos.html>.
- Pressman, R.S. 2005.** *Software Engineering: A Practitioner's Approach*. . 2005.
- 2018.** Programarfacil.com. *Programarfacil.com*. [En línea] 2018. [Citado el: 20 de 02 de 2018.]
<https://programarfacil.com/blog/programacion/los-patrones-de-diseno/>.
- 2001-2018.** python. [En línea] 2001-2018. [Citado el: 01 de 03 de 2017.] <https://www.python.org/>.
- Python Software Foundation (US) . 2001 - 2018.** python. [En línea] 2001 - 2018. [Citado el: 12 de 02 de 2017.] <https://www.python.org/>.
- Python Software Foundation (US). 20001-2018.** phyton. [En línea] 20001-2018. [Citado el: 20 de 03 de 2018.] <https://www.python.org/dev/peps/pep-0008/>.
- Python Software Foundation. 1990-2018.** python. [En línea] 1990-2018. [Citado el: 12 de 04 de 2018.]
<https://docs.python.org/2/library/unittest.html?highlight=uni#module-unittest>.
- Richardson, Leonard. 1998-2018.** Beautiful Soup. [En línea] 1998-2018. [Citado el: 15 de 03 de 2018.]
<https://www.crummy.com/software/BeautifulSoup/>.
- S.P.Ruba, Rani, Ramesh, B y Anusha, M. 2015.** *Evaluation of stemming techniques for text classification*. s.l. : International Journal of Computer Science and Mobile Computing, 2015. 2320-088X.
- Scraping, Web. 2014.** *What is import·io from the user's point of view?* [En línea] 2014. [Citado el: 14 de 11 de 2017.] <http://scraping.pro/what-is-import-io/>.

- Shaw, M. y Garlan, D. 1995.** *“Formulations and Formalisms in Software Architecture”*. s.l. : Springer-Verlag, 1995. Volume 1000—Lecture Notes in Computer Science.
- SOME ASPECTS REGARDING ON TECHNOLOGICAL VIGILA*. **Savescu, Dan. 2014.** 2014.
- Tecnología IALE. 2016.** Vigiale. [En línea] 2016. [Citado el: 21 de 11 de 2017.] <https://www.vigiale.com/>.
- The jQuery Foundation. 2018.** jQuery. [En línea] 2018. <https://jquery.com/>.
- The Use of Web-scraping Software in Searching for Grey Literature*. **Haddaway, Neal R. 2015.** 3, 2015, Vol. 11.
- 2017.** UniMOOC. *UniMOOC*. [En línea] 2017. <https://unimooc.com/>.
- Web Scraping. 2002 - 2014.** Helium Scraper Review. [En línea] 2002 - 2014. [Citado el: 15 de 11 de 2017.] <http://scraping.pro/helium-scraper-review/>.
- **2002 - 2014.** OutWit Hub Review. [En línea] 2002 - 2014. [Citado el: 16 de 11 de 2017.] <http://scraping.pro/outwit-hub-review/>.
- **2002 - 2014.** Software for Web Scraping. [En línea] 2002 - 2014. [Citado el: 05 de 11 de 2017.] <http://scraping.pro/software-for-web-scraping/>.
- **2002 - 2014.** Web Harvy Data Extractor. [En línea] 2002 - 2014. [Citado el: 17 de 11 de 2017.] <http://scraping.pro/sysnucleus-webharvy/>.
- **2002 - 2014.** What is import.io from the user`s point of view. [En línea] 2002 - 2014. [Citado el: 01 de 02 de 2017.] <http://scraping.pro/what-is-import-io/>.
- WebExtractor System. 2018.** Web Data Extractor. Automatically extract specific information from web pages. [En línea] 2018. [Citado el: 15 de 06 de 2017.] <http://www.webextractor.com/help.htm>.
- Wusteman, Judith. 2004.** *RSS: the latest feed*. s.l. : Emerald Group Publishing Limited, 2004.
- Yao, Y.Y. 1995.** *Measuring Retrieval Effectiveness Based on User Preference of Documents*. *Journal of the American Society for Information Science*. 1995.

ANEXOS. CASOS DE PRUEBA DE ACEPTACIÓN

Tabla 56. Caso de Prueba de Aceptación - Autenticar Usuario - Datos Válidos

Caso de prueba de aceptación	
Nombre: Autenticar Usuario - Datos Válidos	
Historia de Usuario: Autenticar Usuario	
Descripción: Se prueba el acceso al sistema con datos (nombre, contraseña) del usuario correctos.	
Precondiciones: Se deben proveer datos (nombre, contraseña) correctos.	
Pasos de ejecución: Entrar a la página de autenticación del sistema. Se introducen los datos correctos de usuario y contraseña.	Se muestra un formulario con los campos de nombre de usuario y contraseña. El Sistema muestra la página principal de sus funcionalidades.

Tabla 57. Caso de Prueba de Aceptación - Autenticarse Usuario - Datos Inválidos

Caso de prueba de aceptación	
Nombre: Autenticarse Usuario - Datos Inválidos	
Historia de Usuario: Autenticarse Usuario	
Descripción: Se prueba el acceso al sistema con datos (nombre, contraseña) incorrectos.	
Precondiciones: Se deben proveer datos (nombre, contraseña) incorrectos.	
Pasos de ejecución: Entrar a la página de autenticación del sistema. Se introducen los datos incorrectos de usuario y contraseña.	Se muestra un formulario con los campos de nombre de usuario y contraseña. El Sistema muestra un mensaje de error: "Error de Inicio".

Tabla 58. Caso de Prueba de Aceptación - Añadir temáticas - Datos Válidos

Caso de prueba de aceptación	
Nombre: Añadir temática - Datos Válidos	
Historia de Usuario: Gestión de temáticas	
Descripción: Se prueba que se añada la temática con los datos (nombre, descripción) correctos.	
Precondiciones: Autenticarse en el sistema.	
Pasos de ejecución: Acceder al módulo Temas. Seleccionar el botón agregar temática. Llenar los datos correspondientes de la temática.	Se muestran los temas agregados y el botón agregar temática. Se muestra un formulario con campos a llenar. Aceptar el formulario se guarda la nueva temática en la base de datos.

Tabla 59. Caso de Prueba de Aceptación - Añadir temáticas - Datos Inválidos

Caso de prueba de aceptación	
Nombre: Añadir temática - Datos Inválidos	
Historia de Usuario: Gestión de temáticas	
Descripción: Se prueba que se añada la temática con datos vacíos.	
Precondiciones: Autenticarse en el sistema.	
Pasos de ejecución: Acceder al módulo Temas. Seleccionar el botón agregar temática. Llenar los datos correspondientes de la temática con datos vacíos.	El sistema muestra todos los temas agregados y el botón agregar temática. Se muestra un formulario con campos a llenar. Al entrar datos incorrectos el sistema se muestra una alerta: "tema vacío" e invalida la entrada de estos datos.

Tabla 60. Caso de Prueba de Aceptación - Mostrar temáticas - Datos Válidos

Caso de prueba de aceptación	
Nombre: Mostrar temática - Datos Válidos	
Historia de Usuario: Gestión de temáticas	
Descripción: Se prueba que se muestre la temática con los datos correctos.	
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener al menos un tema agregado. 	
Pasos de ejecución: Acceder al módulo Temas. Hacer clic sobre uno de los temas que se muestran.	El sistema muestra todos los temas agregados. El sistema muestra los datos que contiene la temática.

Tabla 61. Caso de Prueba de Aceptación - Listar temáticas - Datos Válidos

Caso de prueba de aceptación	
Nombre: Listar temática - Datos Válidos	
Historia de Usuario: Gestión de temáticas	
Descripción: Se prueba que se listen las temáticas con los datos correctos.	
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener al menos un tema agregado. 	
Pasos de ejecución: Acceder al módulo Temas.	El sistema muestra todos los temas agregados.

Tabla 62. Caso de Prueba de Aceptación - Modificar temáticas - Datos Válidos

Caso de prueba de aceptación	
Nombre: Modificar temática - Datos Válidos	
Historia de Usuario: Gestión de temáticas	
Descripción: Se prueba que se modifique la temática con los datos correctos.	
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener al menos un tema agregado. 	
Pasos de ejecución: Acceder al módulo Temas. Hacer clic sobre el botón modificar temática de uno de los temas que se muestran. Se llenan los datos que se piden y se acepta el formulario.	El sistema muestra todos los temas agregados. El sistema muestra un formulario con los datos a modificar.

Tabla 63. Caso de Prueba de Aceptación - Modificar temáticas - Datos Inválidos

Caso de prueba de aceptación	
Nombre: Modificar temática - Datos Inválidos	
Historia de Usuario: Gestión de temáticas	
Descripción: Se prueba que se modifique la temática con los datos incorrectos.	
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener al menos un tema agregado. 	
Pasos de ejecución: Acceder al módulo Temas.	El sistema muestra todos los temas agregados. Se muestra un formulario con campos a llenar.

<p>Seleccionar el botón modificar temática de uno de los temas mostrados.</p> <p>Llenar los datos correspondientes de la temática con datos vacíos.</p>	<p>Al entrar datos vacíos el sistema se muestra una alerta: "tema vacío" e invalida la entrada de estos datos.</p>
---	--

Tabla 64. Caso de Prueba de Aceptación - Eliminar temáticas - Datos Válidos

Caso de prueba de aceptación	
Nombre: Eliminar temática - Datos Válidos	
Historia de Usuario: Gestión de temáticas	
Descripción: Se prueba que se elimine la temática de forma correcta.	
Precondiciones:	
<ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener al menos un tema agregado. 	
Pasos de ejecución:	
Acceder al módulo Temas.	El sistema muestra todos los temas agregados y el botón eliminar temática.
Seleccionar el botón eliminar temática de uno de los temas mostrados.	Se muestra un formulario con todos los temas añadidos.
Se escoge un tema a eliminar y se acepta el formulario.	El sistema elimina la temática de la base de datos.

Tabla 65. Caso de Prueba de Aceptación - Añadir Selectores - Datos Válidos

Caso de prueba de aceptación	
Nombre: Añadir selectores. - Datos Válidos	
Historia de Usuario: Gestión de selectores.	
Descripción: Se prueba que se añada los selectores con los datos correctos.	

Precondiciones: Autenticarse en el sistema.	
Pasos de ejecución:	
Acceder al módulo selectores.	El sistema muestra todos los selectores añadidos a la base de datos y un botón de añadir selector.
Seleccionar botón añadir selector.	El sistema muestra un formulario con los datos necesarios para un selector.
Se llenan los datos y se acepta el formulario.	

Tabla 66. Caso de Prueba de Aceptación - Listar Selectores - Datos Válidos

Caso de prueba de aceptación	
Nombre: Listar selectores. - Datos Válidos	
Historia de Usuario: Gestión de selectores.	
Descripción: Se prueba que se listen los selectores de forma correcta.	
Precondiciones:	
<ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener al menos un selector agregado. 	
Pasos de ejecución:	
Acceder al módulo Selectores.	Se muestran una lista de los selectores que se encuentran registrados.

Tabla 67. Caso de Prueba de Aceptación - Modificar Selectores - Datos Válidos

Caso de prueba de aceptación	
Nombre: Modificar selectores. - Datos Válidos	
Historia de Usuario: Gestión de selectores.	
Descripción: Se prueba que se modifique el selector con los datos correctos.	
Precondiciones:	
<ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener al menos un selector agregado. 	
Pasos de ejecución:	
Acceder al módulo Selectores.	Se muestran una lista de los selectores que se encuentran registrados con el botón modificar.

<p>Presionar el botón modificar del selector que se desea. Se llenan los datos del formulario y se acepta el formulario.</p>	<p>El sistema muestra un formulario con los datos necesarios para los selectores.</p> <p>El sistema modifica los datos del selector.</p>
--	--

Tabla 68. Caso de Prueba de Aceptación - Modificar Selectores - Datos Inválidos

Caso de prueba de aceptación	
Nombre: Modificar selectores. - Datos Inválidos	
Historia de Usuario: Gestión de selectores.	
Descripción: Se prueba que se modifique el selector con los datos incorrectos.	
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener al menos un selector agregado. 	
Pasos de ejecución: Acceder al módulo Selectores. Presionar el botón modificar del selector que se desea. Se dejan en blanco el campo de temática del selector y se acepta el formulario.	<p>Se muestran una lista de los selectores que se encuentran registrados con el botón modificar.</p> <p>El sistema muestra un formulario con los datos necesarios para los selectores.</p> <p>El sistema muestra una alerta: "Tema Vacío" y no modifica el selector.</p>

Tabla 69. Caso de Prueba de Aceptación - Eliminar Selectores - Datos Válidos

Caso de prueba de aceptación	
Nombre: Eliminar selectores. - Datos Válidos	

Historia de Usuario: Gestión de selectores.	
Descripción: Se prueba que se modifique el selector con los datos correctos.	
Precondiciones:	
<ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener al menos un selector agregado. 	
Pasos de ejecución:	
Acceder al módulo Selectores.	Se muestran una lista de los selectores que se encuentran registrados con el botón eliminar selector.
Presionar el botón eliminar del selector que se desea.	Se muestra una alerta: "Está seguro que desea eliminar el selector".
Se acepta la alerta.	El sistema elimina el selector.

Tabla 70. Caso de Prueba de Aceptación - Añadir URL - Datos Válidos

Caso de prueba de aceptación	
Nombre: Añadir URL - Datos Válidos	
Historia de Usuario: Gestión de URL	
Descripción: Se prueba que se puedan añadir de manera manual una URL de forma correcta.	
Precondiciones:	
<ul style="list-style-type: none"> ➤ Autenticarse usuario. 	
Pasos de ejecución:	
Acceder al módulo URLs.	Se muestran todas las URLs añadidas al sistema y el botón agregar URL.
Presionar el botón agregar URL.	El sistema muestra un formulario con los datos necesarios para las URLs
Llenar los datos de la URL que desea agregar y aceptar el formulario	El sistema agrega una URL a la base de datos.

Tabla 71. Caso de Prueba de Aceptación - Añadir URL - Datos Inválidos

Caso de prueba de aceptación	
Nombre: Añadir URL - Datos Inválidos	
Historia de Usuario: Gestión de URL	
Descripción: Se prueba que se puedan añadir de manera manual una URL de forma incorrecta.	
Precondiciones:	
<ul style="list-style-type: none"> ➤ Autenticarse usuario. 	
Pasos de ejecución:	
Acceder al módulo URLs.	Se muestran todas las URLs añadidas al sistema y el botón agregar URL.
Presionar el botón agregar URL.	El sistema muestra un formulario con los datos necesarios para las URLs.
Llenar los datos de la URL que desea agregar con datos de la temática vacía y aceptar el formulario	El sistema muestra una alerta: “temática vacía” y no agrega la URL.

Tabla 72. Caso de Prueba de Aceptación - Añadir URL según los resultados de internet - Datos Válidos

Caso de prueba de aceptación	
Nombre: Añadir URL según los resultados de internet - Datos Válidos	
Historia de Usuario: Gestión de URL	
Descripción: Se prueba que se puedan añadir la URL a partir de una búsqueda en internet forma incorrecta.	
Precondiciones: El sistema o el usuario deben tener un criterio de búsqueda antes de añadir una URL desde internet.	
Pasos de ejecución: Acceder al módulo buscar URL de interés. Se presiona el botón búsqueda. Seleccionar los de interés para el usuario, y presionar el botón añadir resultados.	<p>El sistema muestra un cuadro de búsqueda donde se pondrá el criterio a buscar, junto a un botón de iniciar búsqueda.</p> <p>El sistema lista los resultados encontrados en internet.</p> <p>El sistema añade las direcciones de los sitios seleccionados a la base de datos.</p>

Tabla 73. Caso de Prueba de Aceptación - Listar URL - Datos Válidos

Caso de prueba de aceptación	
Nombre: Listar URL - Datos Válidos	
Historia de Usuario: Gestión de URL	
Descripción: Se prueba que se puedan listar las URLs de forma correcta.	
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse usuario. ➤ Tener al menos una URL agregada. 	
Pasos de ejecución: Acceder al módulo URLs.	Se muestran una lista de las URLs registradas.

Tabla 74. Caso de Prueba de Aceptación - Modificar URL - Datos Válidos

Caso de prueba de aceptación	
Nombre: Modificar URL - Datos Válidos	
Historia de Usuario: Gestión de URL	
Descripción: Se prueba que se pueda modificar la URL de forma correcta.	
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse usuario. ➤ Tener al menos una URL agregada. 	
Pasos de ejecución: Acceder al módulo URLs. Presionar el botón modificar URL. Llenar los datos de la URL que desea modificar y aceptar el formulario.	<p>Se muestran todas las URLs añadidas al sistema y el botón modificar URL.</p> <p>El sistema muestra un formulario con los datos necesarios de la URLs a modificar.</p> <p>El sistema modifica los datos de la URL correspondiente.</p>

Tabla 75. Caso de Prueba de Aceptación - Modificar URL - Datos Inválidos

Caso de prueba de aceptación	
Nombre: Modificar URL - Datos Inválidos	
Historia de Usuario: Gestión de URL	
Descripción: Se prueba que se pueda modificar la URL de forma incorrecta.	
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse usuario. ➤ Tener al menos una URL agregada. 	

<p>Pasos de ejecución:</p> <p>Acceder al módulo URLs.</p> <p>Presionar el botón modificar URL.</p> <p>Llenar los datos de la URL que desea modificar con datos de la dirección URL vacía y aceptar el formulario</p>	<p>Se muestran todas las URLs añadidas al sistema y el botón modificar URL.</p> <p>El sistema muestra un formulario con los datos necesarios para modificar las URLs.</p> <p>El sistema muestra una alerta: “Dirección URL vacía” y no agrega la URL.</p>
---	---

Tabla 76. Caso de Prueba de Aceptación - Eliminar URL - Datos Válidos

Caso de prueba de aceptación	
Nombre: Eliminar URL - Datos Inválidos	
Historia de Usuario: Gestión de URL	
Descripción: Se prueba que se pueda eliminar la URL de forma correcta.	
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse usuario. ➤ Tener al menos una URL agregada. 	
<p>Pasos de ejecución:</p> <p>Acceder al módulo URLs.</p> <p>Presionar el botón eliminar URL correspondiente a dicha URL.</p> <p>Se acepta la alerta.</p>	<p>Se muestran todas las URLs añadidas al sistema y el botón eliminar URL.</p> <p>El sistema muestra una alerta: “Esta segura(o) que desea eliminar la URL”.</p> <p>El sistema elimina la URL seleccionada.</p>

Tabla 77. Caso de Aceptación - Configurar automáticamente el monitoreo - Datos Válidos

Caso de prueba de aceptación	
Nombre: Configurar automáticamente el monitoreo - Datos Válidos	
Historia de Usuario: Configurar automáticamente el monitoreo.	

Descripción: Probar la configuración del monitoreo automático y sus resultados.	
Precondiciones:	
<ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener la fecha y hora a monitorear. 	
Pasos de ejecución:	
Acceder al módulo monitoreo automático	Se muestra una lista de todos los monitoreos automáticos agregados en el sistema y el botón agregar monitoreo.
Presionar el botón de agregar nuevo monitoreo.	El sistema muestra un formulario con los datos a llenar.
Se llenan los datos correspondientes para un nuevo monitoreo y se acepta el formulario.	El sistema registra el nuevo monitoreo en la base de datos.

Tabla 78. Caso de Aceptación - Configurar automáticamente el monitoreo - Datos Inválidos

Caso de prueba de aceptación	
Nombre: Configurar automáticamente el monitoreo - Datos Inválidos	
Historia de Usuario: Configurar automáticamente el monitoreo.	
Descripción: Probar que la configuración del monitoreo automático y sus resultados se realice de manera incorrectamente	
Precondiciones:	
<ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener la fecha y hora a monitorear. 	
Pasos de ejecución:	
Acceder al módulo monitoreo automático.	Se muestra una lista de todos los monitoreos automáticos agregados en el sistema y el botón agregar monitoreo.
Presionar el botón de agregar nuevo monitoreo.	El sistema muestra un formulario con los datos a llenar.
Al llenar los datos correspondientes para un nuevo monitoreo se deja vacío el campo para la fecha y se acepta el formulario.	El sistema muestra una alerta: "Fecha vacía" y no se agrega el nuevo monitoreo.

Tabla 79. Caso de Aceptación - Mostrar trazas de las búsquedas realizadas - Datos Válidos

Caso de prueba de aceptación

Nombre: Mostrar trazas de las búsquedas realizadas - Datos Válidos
Historia de Usuario: Listar el registro de la búsqueda y el número de URLs analizadas.
Descripción: Probar que se muestren los tiempos de inicio de las búsquedas realizadas.
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Haber configurado o realizado manualmente una búsqueda.
Pasos de ejecución: Al acceder al módulo búsquedas se muestra un registro de todas las búsquedas que se realizaron con el sistema.

Tabla 80. Caso de Aceptación - Listar resultados de las búsquedas - Datos Válidos

Caso de prueba de aceptación	
Nombre: Listar resultados de las búsquedas - Datos Válidos	
Historia de Usuario: Gestionar los resultados de las búsquedas	
Descripción: Probar que se muestren los resultados de las búsquedas en las URL por temática.	
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener al menos un resultado de la búsqueda. 	
Pasos de ejecución: Acceder al módulo Temas. Acceder a un tema específico. Acceder a una dirección específica de las mostradas.	El sistema muestra todos los temas agregados y el botón agregar temática. El sistema muestra todas las direcciones URL relacionadas con ese tema. El sistema muestra todos los resultados de las búsquedas realizadas para esa dirección.

Tabla 81. Caso de Aceptación - Eliminar resultados de las búsquedas - Datos Válidos

Caso de prueba de aceptación
Nombre: Eliminar resultados de las búsquedas - Datos Válidos

Historia de Usuario: Gestionar los resultados de las búsquedas	
Descripción: Probar que se muestren los resultados de las búsquedas en las URL por temática.	
Precondiciones:	
<ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener al menos un resultado de la búsqueda. 	
Pasos de ejecución:	
Listar resultados de las búsquedas.	El sistema muestra el listado donde cada bloque de información contiene un botón eliminar resultado.
Presionar botón eliminar resultado.	El sistema muestra una alerta: "Esta seguro que desea eliminar el resultado"
Se acepta la alerta del sistema	El sistema elimina el resultado de la base de datos.

Tabla 82. Caso de Aceptación - Modificar resultados de las búsquedas - Datos Válidos

Caso de prueba de aceptación	
Nombre: Modificar resultados de las búsquedas - Datos Válidos	
Historia de Usuario: Gestionar los resultados de las búsquedas	
Descripción: Probar que se muestren los resultados de las búsquedas en las URL por temática.	
Precondiciones:	
<ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener al menos un resultado de la búsqueda. 	
Pasos de ejecución:	
Una vez listados los resultados de la búsqueda se muestran para cada uno la clasificación fusificada que da el sistema del mismo.	
El usuario puede cambiar esa clasificación escogiendo dentro de las tres posibles.	
Presionar el botón modificar del resultado.	

Tabla 83. Caso de Aceptación - Generar documento PDF con resultados seleccionados - Datos Válidos

Caso de prueba de aceptación

Nombre: Generar documento PDF con resultados seleccionados - Datos Válidos	
Historia de Usuario: Generar documento PDF con resultados seleccionados	
Descripción: Probar que se genere un documento PDF con los resultados relevantes de gestión de las búsquedas.	
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener información recopilada de las búsquedas y marcada para ser añadidas. 	
Pasos de ejecución: Acceder al módulo temas. Presionar icono de generar PDF.	<p>El sistema muestra todos los temas almacenados en la base de datos además de un icono donde se puede generar PDF.</p> <p>El sistema generará un documento en formato .pdf con todos los resultados del tema marcados para ser añadidos.</p> <p>Y muestra una alerta: "Se creó el nuevo documento PDF"</p>

Tabla 84. Caso de Aceptación - Configurar envío de informes por correo - Datos Válidos

Caso de prueba de aceptación	
Nombre: Configurar envío de informes por correo - Datos Válidos	
Historia de Usuario: Gestionar envío de informes por correo	
Descripción: Probar el envío de correo a una dirección específica con los resultados de la búsqueda.	
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ Tener una dirección de correo válida. ➤ Tener información recopilada y marcada para ser añadida. 	
Pasos de ejecución: Acceder al módulo temas. Se llenan los datos del formulario y se presiona el botón enviar correo.	<p>Se muestra el botón enviar correo.</p> <p>El sistema muestra un formulario con los datos para el correo a enviar.</p> <p>El correo se enviará con los resultados marcados para el envío.</p>

Tabla 85. Caso de Aceptación - Configurar envío de informes por correo - Datos Inválidos

Caso de prueba de aceptación	
Nombre: Configurar envío de informes por correo - Datos Inválidos	
Historia de Usuario: Configurar envío de informes por correo	
Descripción: Probar el envío de correo a una dirección específica con los resultados de la búsqueda se realice de forma incorrecta.	
Precondiciones: <ul style="list-style-type: none"> ➤ Autenticarse en el sistema. ➤ No tener una dirección de correo válida. ➤ No tener información recopilada. 	
Pasos de ejecución: Acceder al módulo temas. Se llenan los datos del formulario dejando la dirección de correo incoherente, se presiona el botón enviar correo.	Se muestra el botón enviar correo. El sistema muestra un formulario con los datos para el correo a enviar. El sistema muestra una alerta: "dirección de correo inválida". Y no envía el correo.