

Universidad de las Ciencias Informáticas

Facultad 4



Análisis y diseño de un Sistema de Vigilancia Tecnológica para la tecnología educativa

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

Autores:

Yuniel Perdomo Fernández
Arianna Gámez Columbié

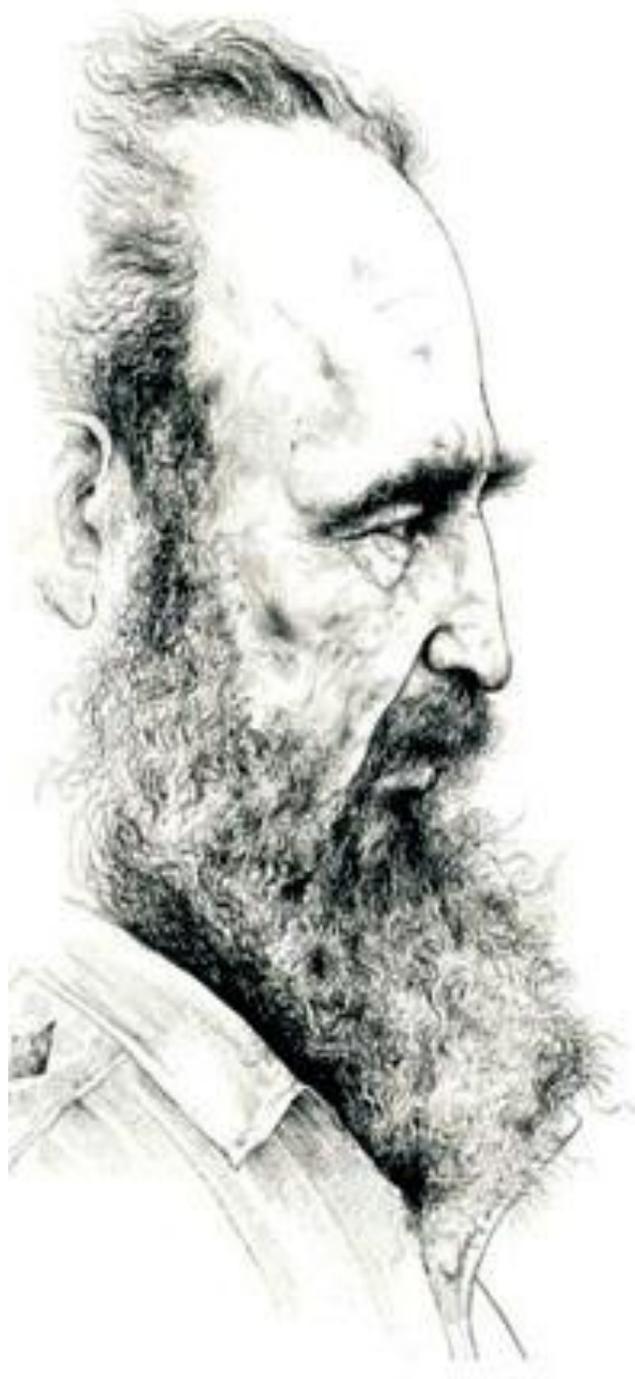
Tutores:

Lic. Yenieris Moyares Norchales
Ing. Maikel Aparicio Reytor

Co-tutor:

Ing. Yunior Orosa Velázquez

La Habana, junio de 2012
“Año 54 de la Revolución”



"...Creo que un importante trabajo en el área científica de nuestros investigadores, de nuestros profesionales, es recoger información, recoger experiencia de lo que se hace en el mundo, y me parece que en eso hemos estado atrasados..."

Fidel Castro Ruz

*15 de Enero de 1997
Día de la Ciencia Cubana*

A large, stylized handwritten signature of Fidel Castro Ruz, written in black ink. The signature is highly cursive and expressive, with a prominent horizontal stroke at the top and bottom.

DEDICATORIA

A mis padres por todo el amor y la confianza depositada en mí durante toda la carrera, por estar en cada momento cuando más los necesité, por ser mi motivación, por darme siempre todo su apoyo incondicional, por ser ejemplos y guías; por el orgullo que sienten hoy. A mis abuelitos y abuelita, que no están presentes hoy pero que estoy segura les hubiese gustado mucho verme graduada. A mi abu Blasita que ha estado orando por mí todo este tiempo.

Arianna

A mi padres por ser lo más grande y hermoso que tengo, mi razón de ser, mi inspiración. Mi ejemplo a seguir; por depositar en mí toda la confianza durante mi carrera, por su apoyo incondicional en todo momento, en fin porque sin ellos hubiera sido difícil llegar hasta aquí.

A mis hermanos; Yoel, mi ejemplo a seguir, el primer informático de la familia que siempre estuvo enviando sus correos alentándome y dándome cursos de responsabilidad. A Landi mi hermanito del alma para que siga mi ejemplo y mañana sea un profesional, no importa cuántas trabas te ponga la vida en el camino siempre lograrás lo que te propongas y también a Ñeco el más pequeñín y serio de todos.

A mis abuelos y abuelas que hoy deben estar rezando mucho por mí, mi abuelo Pedro que debe estar ahí muy cerca de mí como antes cuidando cada paso que doy.

A mi tía Nelly el faro de la familia que se estará muy orgullosa de mí.

Yuniel

Agradecimientos

AGRADECIMIENTOS

A Dios, porque sin su bendición nada puede ser posible.

Al Comandante en Jefe Fidel Castro Ruz por haberme dado la posibilidad de estudiar y hacerme ingeniera en una Universidad de excelencia.

A mi preciosa madre por darme siempre las fuerzas necesarias para continuar adelante, por estar ahí en todos los momentos de mi vida.

A mi grandioso padre por ser mi orgullo y ejemplo a seguir, por inculcarme todo lo que soy, por dedicar todas sus horas de trabajo a mi persona.

A mi linda familia que siempre ha estado a mi lado apoyándome

A mi fiel amigo, mi gran amor y compañero de tesis por darme todo su apoyo incondicional y su amor, gracias por llegar en el momento que más lo necesitaba y compartir maravillosos días conmigo.

Arianna

A mi novia, amiga, mi niña mimada que siempre estuvo ahí para mí, en los momentos malos y en lo buenos, la que me dio fuerzas cada día para seguir, para levantarme, mi mayor inspiración para terminar esta tesis. A ella debo todo lo que soy hoy.

A mi familia aquí en la habana por el apoyo que me dieron durante estos 5 años. Mi tia Magalís y Marbelis. A mis primas Yumi, Yanelkis, Patricia; a Eliades, Osmanis que siempre estuvo ahí para sacarme de los aprietos.

A toda mi familia por el apoyo y la confianza depositada en mí.

A mis compañeros de estudios durante toda la carrera Mayliuvis, Miroslava, Yolaida, Rajiv, Nesto, Roberto, Lisandrita, Edisnay, Jose Carlos, Diosbel, Agustin, Anabel, Iraisi, Kenier, Yuleisy, Alejandro, Milena, Yoandi, a todos los que se me pudieran quedar y que compartieron buenos y malos momentos en la universidad.

A mis amigos Carlos (Charlie) y Mario por su apoyo y dedicación. A los de Baracoa Amelia, Daikel, Alejandro porque siempre están ahí y van conmigo de la mano.

Yuniel

Agradecimientos

A nuestros tutores Maiquel y Yeni que nos dedicaron muchas horas de sueño revisando lo más mínimo, su apoyo fue crucial para el éxito de la tesis.

A Yunior (cotutor), con sus comentarios obscenos, que vió salir el sol junto a nosotros revisando cada detalle del documento.

A Leonardo, Linet y Toni que cada uno puso su granito de arena para que todo saliera bien. Yusdel compañero de la vieja FEU hoy amigo que nos ayudó desde que estábamos en Venezuela con el tema de tesis y así en innumerables ocasiones que siempre estuvo ahí.

Arianna y Yuniel

Declaración de autoría

DECLARACIÓN DE AUTORÍA

Declaramos que somos autores únicos de este trabajo y autorizamos al Centro de Tecnologías para la Formación (FORTES) de la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

Para que así conste firmo la presente a los 28 días del mes de junio del año 2012.

Autores:

Arianna Gámez Columbié

Yuniel Perdomo Fernández

Tutores:

Lic. Yenieris Moyares Norchales

Ing. Maikel Aparicio Reytor

Co-tutor:

Ing. Yunior Orosa Velázquez

RESUMEN

La vigilancia tecnológica es una práctica indispensable e innovadora para la adquisición y gestión de la información convirtiéndola en conocimiento aplicada a proyectos de desarrollo de productos, así mismo se utiliza como recurso en las empresas para minimizar el riesgo de las decisiones, centrar los objetivos, detectar nuevas ideas, conocer el entorno, anticiparse al cambio y desarrollar productos con calidad y exclusividad. El presente trabajo tiene como objetivo elaborar un sistema de vigilancia tecnológica, el cual contribuirá a prestar servicios de inteligencia de manera ágil para la toma de decisiones en el grupo de Gestión de la Información y el Conocimiento (GIC) del Centro de Tecnologías para la Formación (FORTES).

PALABRAS CLAVE

Gestión de la información, sistemas de vigilancia, vigilancia tecnológica.

Índice de contenidos

INTRODUCCIÓN.....	1
1. CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA	5
1.1 La vigilancia. Funciones y ventajas para la competitividad.....	5
1.1.1 Conceptos de vigilancia	5
1.1.2 Necesidad de la vigilancia.....	6
1.1.3 Tipos de vigilancia.....	6
1.1.4 Tipo de vigilancia seleccionada.....	8
1.2 Vigilancia tecnológica (VT)	10
1.2.1 Conceptos	10
1.2.2 Beneficios de la VT	10
1.2.3 Fuentes de información.....	13
1.3 Herramientas para la VT.....	14
1.3.1 Principales tendencias de las herramientas de VT	14
1.3.2 Herramientas internacionales.....	15
Tetralogie	15
Matheo Software	16
Vigtech.....	16
Goldfire	17
1.3.3 Herramientas nacionales	17
DataSOMining (Instituto Finlay, Universidad Nacional Autónoma de México)	17
BioMundiPatent (Consultoría Biomundi/IDICT).....	18
ProIntec (Universidad de Pinar del Río).....	18
CiteLivePro 2.0 (Centro de Bioplantas, Universidad de Ciego de Ávila)	18
InfoCam v3.0	18
Mimosa (OCPI).....	19
Sistema automatizado para la vigilancia de patentes SiVigPat (CENIC)	19
ToolInf (Consultoría Biomundi/IDICT)	19
Centro de Información y Gestión Tecnológica (CIGET), Villa Clara.....	20

Índice de contenidos

Valoración sobre las herramientas	20
1.4 Bibliometría y cienciometría.....	21
1.5 Tecnologías para el diseño.....	22
1.5.1 Lenguaje de modelado UML 2.1	22
1.5.2 Herramientas de Ingeniería de Software Asistida por Computadora (CASE)	22
1.5.2.1 Visual Paradigm for UML 8.0	23
1.5.2.2 Rational Rose Enterprise 7.0	23
Consideraciones de las Herramientas CASE.....	24
1.5.3 Metodologías de desarrollo de software	25
1.5.3.1 Rational Unified Process (RUP) [25].....	25
1.5.3.2 Extreme Programming (XP) [25]	27
Metodología seleccionada.....	30
Conclusiones parciales	31
2. CAPÍTULO 2. PROPUESTA DEL SISTEMA	32
2.1 Descripción del sistema de vigilancia tecnológica	32
2.1.1 Adquisición de la información.....	33
2.1.2 Análisis y procesamiento de la información	34
2.1.2.1 Extracción de características e información útil de los documentos	35
2.1.2.2 Análisis de estadísticas descriptivas y multivariadas.....	35
2.1.2.3 Análisis de redes sociales (ARS)	36
2.1.3 Difusión de la información	37
2.2 Propuesta de Sistema Gestor de Contenidos (CMS)	38
2.3 Arquitectura	40
2.4 Propuesta de tecnologías para la implementación del sistema	41
2.4.1 Lenguajes de programación.....	41
2.4.1.1 Lenguajes del lado del cliente. [30].....	42
2.4.1.2 Lenguajes del lado del servidor	43
2.4.2 Gestor de base de datos	44
Valoración de los gestores de bases de dato	46

Índice de contenidos

2.4.3	Servidores web	46
	Consideración de las tecnologías propuestas para el desarrollo.....	47
2.5	Roles para el sistema de vigilancia.....	47
2.6	Modelo de dominio	48
2.7	Especificación de Requisitos	49
2.7.1	Requisitos Funcionales	50
2.7.2	Requisitos no funcionales	51
2.8	Modelo de Casos de Uso del Sistema.....	53
2.8.1	Patrones de Casos de Uso	53
2.8.2	Descripción de casos de uso del sistema	55
2.9	Análisis y diseño	56
2.9.1	Modelo de Análisis	56
2.9.1.1	Diagramas de interacción del análisis	58
2.9.2	Modelo de diseño.....	58
2.9.2.1	Patrones de diseño.....	59
2.9.2.2	Diagrama de clases del diseño.....	60
	Conclusiones Parciales	61
3.	CAPÍTULO 3. VALIDACIÓN DE LA PROPUESTA.....	62
3.1	Métricas	62
3.2	Métrica para la Calidad de la Especificación de los Requisitos.	63
3.3	Métricas para evaluar el diseño	67
	Conclusiones parciales	68
	CONCLUSIONES GENERALES	69
	RECOMENDACIONES.....	70
	REFERENCIAS BIBLIOGRÁFICAS	71
	BIBLIOGRAFÍA.....	73
	GLOSARIO DE TÉRMINOS	76
	ANEXOS.....	78
	Anexo 1. Conceptos abordados por teóricos de vigilancia tecnológica.....	78
	Anexo 2. Comparación de las herramientas Matheo software.....	79

Índice de contenidos

Anexo 3. Descripciones de los casos de uso.....	79
Anexo 4. Diagramas de Clases del Análisis.....	87
Anexo 5. Diagramas de Colaboración.....	91
Anexo 6. Diagramas de Clases del Diseño.....	102
Anexo 7. Resultados de la validación de los requerimientos.....	109
Anexo 7.1. Corrección.....	109
Anexo 7.2. Compleción.....	110
Anexo 7.3. Comprensión.....	110
Anexo 7.4. Consistencia interna.....	111
Anexo 7.5. Consistencia externa.....	111
Anexo 8. Cálculo del índice de Jaccard.....	112
Anexo 9. Fórmula de Coeficiente de DICE.....	113
Anexo 10. Histogramas.....	113
Anexo 11. Aval de calidad.....	114

Índice de figuras

Figura 1. Clasificación de la vigilancia de Porter y la relación entre ellas.	7
Figura 2. Problemáticas-Alternativas detectadas en el Centro FORTES.	8
Figura 3. Fases del proceso de vigilancia tecnológica (I).	9
Figura 4. Fases del proceso de vigilancia tecnológica (II).	9
Figura 5. Principales tendencias en las herramientas de vigilancia e inteligencia.	14
Figura 6. RUP Costo del cambio.	25
Figura 7. Flujos de trabajo de RUP.	26
Figura 8. XP Costo del cambio.	28
Figura 9. Partes de XP.	30
Figura 10. Esquema de sistema de vigilancia tecnológica para el grupo GIC.	32
Figura 11. Proceso de adquisición de la información.	33
Figura 12. Proceso de análisis de la información manual.	34
Figura 13. Proceso de análisis de la información automatizada.	35
Figura 14. Difusión de la información del sistema.	37
Figura 15. Arquitectura de software del CMS Drupal.	41
Figura 16. Esquema de las partes que componen un elemento HTML.	43
Figura 17. Modelo de dominio del sistema de vigilancia tecnológica.	49
Figura 18. Diagrama de caso de uso del sistema.	55
Figura 19. Diagrama de clase del análisis Buscar coocurrencia de palabras.	57
Figura 20. Diagrama de colaboración Buscar coocurrencia de palabras.	58
Figura 21. Diagrama de clase diseño Buscar coocurrencia de palabras.	61
Figura 22. Diagrama de clase del análisis Gestionar fuentes.	87
Figura 23. Diagrama de clase del análisis Gestionar autor.	87
Figura 24. Diagrama de clase del análisis Normalizar palabras claves.	88
Figura 25. Diagrama de clase del análisis Contar palabras.	88
Figura 26. Diagrama de clase del análisis Contar palabras claves por año.	89
Figura 27. Diagrama de clase del análisis Contar autores por año.	89
Figura 28. Diagrama de clase del análisis Modelar relaciones sociales.	90
Figura 29. Diagrama de colaboración Insertar Fuentes.	91
Figura 30. Diagrama de colaboración Modificar Fuentes.	92
Figura 31. Diagrama de colaboración Eliminar fuentes.	93
Figura 32. Diagrama de colaboración Insertar autor.	94
Figura 33. Diagrama de colaboración Modificar autor.	95
Figura 34. Diagrama de colaboración Eliminar autor.	96
Figura 35. Diagrama de colaboración Normalizar palabras claves.	97
Figura 36. Diagrama de colaboración Contar palabra.	98
Figura 37. Diagrama de colaboración Contar palabras claves por año.	99
Figura 38. Diagrama de colaboración contar autor palabra clave y año.	100
Figura 39. Diagrama de colaboración Modelar relaciones sociales.	101
Figura 40. Diagrama de clase diseño Gestionar fuentes.	102
Figura 41. Diagrama de clase diseño Gestionar autor.	103
Figura 42. Diagrama de clase diseño Normalizar palabras claves.	104
Figura 43. Diagrama de clase diseño Contar palabra.	105
Figura 44. Diagrama de clase diseño Contar palabras claves por año.	106
Figura 45. Diagrama de clase diseño Contar autores por palabra clave y año. ...	107
Figura 46. Diagrama de clase diseño Modelar relaciones sociales.	108

INTRODUCCIÓN

En el mundo de hoy el desarrollo tecnológico no pudiera subsistir de manera efectiva sin la información, representando esto representa un elemento crucial para el éxito de cualquier organización, asegurando su rentabilidad, supervivencia y evolución.

Los constantes cambios en las instituciones conllevan a que permanezcan informadas de las tendencias actuales. El éxito de las empresas de Investigación Desarrollo e Innovación (I+D+i), ya no depende de los recursos tangibles con que cuenta, sino que está influenciado por factores externos a la misma relacionados con la búsqueda de información valiosa para la empresa tales como: tendencias actuales, principales proveedores, amenazas, oportunidades, socios o competidores, regulaciones, entre otras.

“En este entorno, las empresas que no formalizan y explotan sistemáticamente las informaciones de valor se ven desbordadas, no pueden gestionar todas las informaciones que reciben, ignoran informaciones de alto valor, su proceso de toma de decisiones se ralentiza y, en resumen, pierden oportunidades de negocio.” [1]

La explotación de la información debe analizarse en dos dimensiones: *“disponer de información sobre el entorno antes que los competidores y desempeñar nuevas armas competitivas a partir del desarrollo y aprovechamiento de la información interna y su transformación en el conocimiento de la organización”*. [2]

“Se impone la necesidad de contar con adecuadas infraestructuras de comunicación y sistemas de información que permitan gestionar la abundante pero dispersa información del entorno y contribuya a la toma de decisiones y esfuerzos de planificación” [3], siendo aquí donde la vigilancia tecnológica (VT) y la inteligencia competitiva juegan un papel protagónico.”

Según la norma española UNE¹ 166006:2006 AENOR², *“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”*. [4]

¹ Una Norma Española.

² AENOR: Asociación Española de Normalización y Certificación (<https://www.aenor.es/>).

Introducción

Diversas instituciones a nivel nacional implementan la actividad de vigilancia. Ejemplo de ello, el Observatorio tecnológico de la Ciudad Universitaria José Antonio Echeverría (CUJAE), el Instituto de Información Científica y Tecnológica (IDICT) con filiales en todas las provincias que hoy son Centros de Información y Gestión Tecnológica (CIGET). Instituciones como las consultorías BIOMUNDI, DELFOS y AVANTE están a la vanguardia en la vigilancia. Brindan servicios enfocados a la vigilancia comercial, competitiva, del entorno y tecnológica que han apoyado a los directivos en la toma de decisiones.

La Universidad de las Ciencias Informáticas (UCI) es una *“Universidad innovadora de excelencia científica, académica y productiva que forma de manera continua profesionales integrales comprometidos con la patria, soporte de la informatización del país y la competitividad internacional de la industria cubana del software.”* [5]

Esta institución educativa dentro de su estructura organizacional cuenta con el Centro de Tecnologías para la Formación (FORTES) que tiene como misión: *“desarrollar tecnologías que permitan ofrecer servicios y productos para la implementación de soluciones de formación aplicando las Tecnologías de la Información y las Comunicaciones, a todo tipo de instituciones con diferentes modelos de formación y condiciones tecnológicas, garantizando la calidad de las soluciones y la formación de los recursos humanos a partir de investigaciones que combinen los elementos pedagógicos y tecnológicos más avanzados, integrando así los procesos de formación, producción e investigación.”* [6]

El Grupo de Gestión de Información y el Conocimiento (GIC), pertenece al Departamento de Implantación y Soporte Técnico (DIST) del centro FORTES y tiene como misión brindar servicios y productos de información de valor añadido a los usuarios de esta entidad. Conocer las líneas de investigación, el trabajo de los competidores y las empresas líderes de un producto determinado son algunos de los principales objetivos de la vigilancia tecnológica. Partiendo de esta afirmación el grupo GIC está enfocado en la práctica de la actividad de la vigilancia, pues ayuda a mantener actualizado a las diferentes áreas temáticas de la entidad.

El grupo GIC para brindar servicios y productos de información, como resultados de la vigilancia tecnológica, tiene que monitorear manualmente un gran número de fuentes de información. En la actualidad se genera más información de la que el Grupo es capaz de procesar, por lo que la gestión de conocimiento se realiza con dilación y se obvia información que no se descubre mediante el análisis realizado por los gestores del Grupo.

Además de esto no existe un espacio virtual donde se puedan debatir y difundir estos resultados.

En base a la situación anteriormente planteada, se define el siguiente **problema a resolver**:

¿Cómo contribuir a que el grupo GIC desarrolle un sistema que agilice la captación, procesamiento y socialización de la información con un enfoque estratégico?

Para llevar a cabo la investigación se ha definido como **objeto de estudio**: Sistemas de vigilancia tecnológica.

El **campo de acción**: Sistema de vigilancia tecnológica para la tecnología educativa.

En el presente trabajo se formuló la siguiente **idea a defender**: El análisis y diseño de un Sistema de Vigilancia Tecnológica para la tecnología educativa contribuye al desarrollo de un sistema que agilice la captación, procesamiento y socialización de la información para el grupo GIC con un enfoque estratégico.

Siendo el **objetivo general**: Elaborar el análisis y diseño de un sistema de vigilancia tecnológica para al grupo GIC que permita captar, procesar y socializar información.

Tareas investigativas:

- Sistematización de los referentes teóricos metodológicos que sustentan la vigilancia tecnológica.
- Análisis de técnicas de exploración de datos.
- Análisis de las fases del proceso de vigilancia tecnológica.
- Descripción de las fases a automatizar.
- Descripción de los requisitos funcionales y no funcionales del sistema.
- Modelación del sistema de vigilancia.
- Selección de métricas a aplicar para la validación de la propuesta.

La investigación tendrá como resultado el análisis y diseño de un sistema de vigilancia tecnológica para el grupo GIC del centro FORTES.

Se utilizarán los siguientes **métodos de investigación científicos**:

- Teóricos:

Introducción

- ✓ Analítico-Sintético: para analizar conceptos, tecnologías, metodologías y modelos con el objetivo de sintetizar los elementos fundamentales que sustentan la investigación.
 - ✓ Análisis-Histórico-Lógico: posibilitó hacer un diagnóstico de la historicidad de la evolución del problema.
 - ✓ Modelado: para definir gráficamente el sistema.
 - ✓ Análisis documental: en el estudio de fuentes de información como revistas, libros, artículos, tesis con la finalidad de sustentar la investigación.
- Empíricos
- ✓ Encuesta: para validar la especificación de los requisitos de acuerdo a la interpretación de los revisores encuestados.

El trabajo está estructurado en tres capítulos, tal y como se describe a continuación:

CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA. En este capítulo se expone la fundamentación teórica del tema. Se abordan conceptos asociados a la vigilancia tecnológica, se describen herramientas de vigilancia tecnológica que se han desarrollado. Además se seleccionan las metodologías y herramientas que se van a utilizar para el diseño del sistema.

CAPITULO 2: PROPUESTA DEL SISTEMA. En este capítulo se detallan los procesos que se llevarán a cabo para la implantación del sistema de VT, describe el negocio, se enumeran los requisitos del sistema para tener un mejor entendimiento de lo que se quiere concebir. Se definen los actores del sistema y se describen los casos de uso. Además, se realiza el análisis y el diseño del sistema.

CAPITULO 3: VALIDACIÓN DE LA PROPUESTA. Se realiza validación del análisis y diseño mediante métricas.

Capítulo 1. Fundamentación Teórica

1. CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

Introducción.

Con la vigilancia tecnológica, se descubren fuentes de información, se extrae información relevante sobre tendencias tecnológicas, novedades, invenciones, potenciales socios o competidores, aplicaciones tecnológicas emergentes, además se contemplan aspectos regulatorios y de mercado que pueden condicionar el éxito de una innovación tecnológica. Todos los elementos planteados bien codificados y analizados brindan a un decisor la posibilidad de trazar estrategias tecnológicas para la empresa, minimizando la incertidumbre.

En este capítulo se plantean los aspectos teóricos de vital importancia para la investigación; incluye una descripción del objeto de estudio, con la finalidad de comprender el proceso de VT que es la base de esta investigación. Se analizarán los diferentes tipos de vigilancia tecnológica, sistemas similares con el objetivo de profundizar más en el tema que se investiga, la necesidad de la vigilancia para la institución. Se exponen los lenguajes, servidores, arquitecturas y patrones que se proponen en el proceso de desarrollo del software.

1.1 La vigilancia. Funciones y ventajas para la competitividad.

1.1.1 Conceptos de vigilancia

La vigilancia de forma general ha sido conceptualizada por varios autores, a continuación alguna de estas definiciones.

Según González Limas la vigilancia, de manera general, puede definirse como: *“el esfuerzo sistemático y organizado por la empresa para la observación, captación, análisis, difusión precisa y recuperación de información sobre los hechos del entorno económico, social o comercial, relevantes para la misma por poder implicar una oportunidad o amenaza para ésta, con objeto de poder tomar decisiones con menor riesgo y poder anticiparse a los cambios”*. [7]

Rouach plantea que: *“es el arte de descubrir, recolectar, tratar, almacenar informaciones y señales pertinentes, débiles y fuertes que permitirán orientar el futuro y proteger el presente y el futuro de los ataques de la competencia”*. [8]

En el trabajo de Fernando Palop y José M. Vicente se define la vigilancia como: *“el esfuerzo sistemático y organizado por la empresa de observación, captación, análisis,*

Capítulo 1. Fundamentación Teórica

difusión precisa y recuperación de información sobre los hechos del entorno económico, tecnológico, social o comercial, relevantes para la misma por poder implicar una oportunidad u amenaza para ésta”. [9]

Los autores definen la vigilancia como el esfuerzo que realiza la institución para alertar sobre oportunidades o amenazas a través de la planificación, captación, procesamiento y difusión de la información sobre el ámbito interno o externo que pueda afectar la empresa.

1.1.2 Necesidad de la vigilancia

Las instituciones de I+D+i siempre han realizado sus búsquedas de forma “tradicional” accediendo a algunos sitios especializados, revistas y eventos científicos, entre otros. Este tipo de vigilancia no es suficiente, la institución debe estar al tanto en todo momento de los movimientos tecnológicos que suceden en el entorno, oportunidades, amenazas y tecnologías emergentes para conseguir ventaja competitiva y evitar sorpresas desagradables.

El desarrollo de las tecnologías ha propiciado que se publiquen numerosos artículos científicos y se registren nuevas patentes diariamente, por lo que se hace necesario contar con un método adecuado que facilite la sistematización del proceso de vigilancia, buscando esencialmente mantenerse al tanto de los temas a fines a la institución. La vigilancia tecnológica constituye una herramienta de alto impacto, que contribuye al mejoramiento de la organización y al sostenimiento de su liderazgo en el mercado.

Según Morin “**vigilar** es una de las seis funciones necesarias para una buena gestión tecnológica, en la que se trata de seguir la evolución de los nuevos adelantos científico-técnicos.” [10]

1.1.3 Tipos de vigilancia

A partir del modelo de Michael Porter de fuerzas que caracterizan la posición competitiva de la empresa, Martinet, B. y Ribault, J (1989) hablan de **cuatro tipos de vigilancia** (Ver figura # 1):

Capítulo 1. Fundamentación Teórica



Figura 1. Clasificación de la vigilancia de Porter y la relación entre ellas.

Fuente: Laura Arribas Martínez. Diseño de un sistema de Vigilancia Tecnológica e Implantación en la Fundación ASCAMM. Abril 2008, Barcelona.

Vigilancia tecnológica o centrada en el seguimiento de los avances del estado de la técnica y en particular de la tecnología y de las oportunidades / amenazas que genera.

Vigilancia competitiva, implica un análisis y seguimiento de los competidores actuales, potenciales y de aquellos con producto sustitutivo.

Vigilancia comercial, dedica la atención sobre los clientes y proveedores.

Vigilancia del entorno, centra la observación sobre el conjunto de aspectos sociales, legales, medioambientales y culturales, que configuran el marco de la competencia.

“Para Jakobiak, F. (1991 y 1992), responsable de vigilancia del grupo petroquímico francés Elf-Atochem, la vigilancia debe articularse sobre unos factores críticos que varían en función de la estrategia y posición de la empresa. Estos factores corresponden a aquellos factores críticos de competitividad a los que cualquier cambio en el entorno de la empresa puede afectarles de forma relevante.” [9]

Capítulo 1. Fundamentación Teórica

Por tal motivo es importante para una buena vigilancia identificar los factores críticos que afectarían directamente las actividades del centro FORTES constituyendo una prioridad a seguir.

1.1.4 Tipo de vigilancia seleccionada

Dado que el objetivo principal del sistema de vigilancia es proveer información contrastada y analizada para aumentar la competitividad de las instituciones donde estén implantados, se realizará la vigilancia en todas aquellas áreas que afecten directamente al entorno competitivo del sector.

En el Centro ya se ha realizado vigilancia tecnológica aunque no de forma sistematizada. Por lo tanto, la disyuntiva que se presenta es la siguiente:

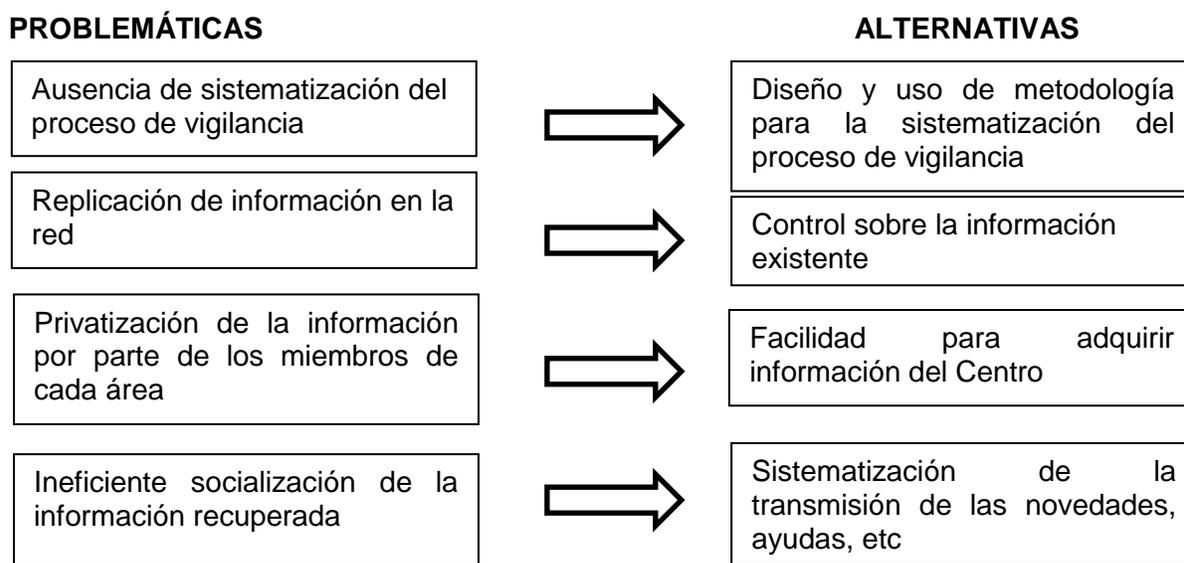


Figura 2. Problemáticas-Alternativas detectadas en el Centro FORTES.

Fuente: Elaboración propia

A partir de la problemática existente y la propuesta de proceso de vigilancia (Ver figuras 3 y 4) que Linnet Ivet Bello Leyva plantea en el 2010 en su tesis de pre-grado, se pretende diseñar un sistema de vigilancia tecnológica que sea implantado para el grupo GIC.

Capítulo 1. Fundamentación Teórica



Figura 3. Fases del proceso de vigilancia tecnológica (I).

Fuente: Linnet Ivet Leyva Bello. Propuesta de proceso de Vigilancia Tecnológica para la implementación de un Observatorio de Tecnología Educativa.2010, UCI.

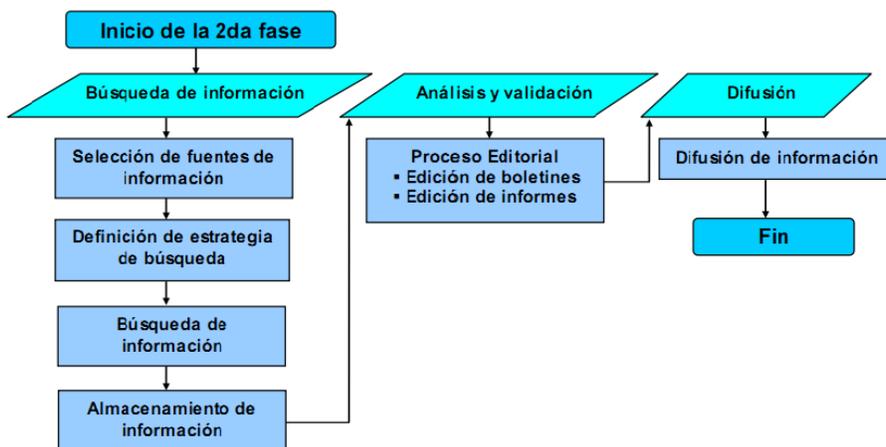


Figura 4. Fases del proceso de vigilancia tecnológica (II).

Fuente: Linnet Ivet Leyva Bello. Propuesta de proceso de Vigilancia Tecnológica para la implementación de un Observatorio de Tecnología Educativa.2010, UCI.

Esta es la base de la que se ha partido para diseñar el sistema de vigilancia tecnológica, adaptando cada etapa a las necesidades propias del grupo GIC. Aun así, prácticamente la totalidad de las acciones a emprender para implantar el sistema de vigilancia son aplicables a cualquier entidad que necesite los beneficios de un sistema similar.

Capítulo 1. Fundamentación Teórica

1.2 Vigilancia tecnológica (VT)

1.2.1 Conceptos

Diversos son los autores teóricos que han conceptualizado el proceso de vigilancia tecnológica en el devenir del tiempo y de los avances tecnológicos. Ejemplos de ellos son: Jakobiak (1992) [8], Lesca (1994) [8], Werner & Degoul (1994) [11], Martinet y Martí, (1995) [8], Escorsa y Maspons (2001) [12], Morcillo (2003) [13] entre otros, para ver estas definiciones consultar el Anexo 1.

Según la Norma Experimental Francesa: XPX50-053. 1998. AFNOR la vigilancia tecnológica es: "*Conjunto de acciones coordinadas de búsqueda, procesamiento y distribución con miras a su explotación, información útil a los actores económicos.*" [14]

Según la norma española UNE 166006:2006 AENOR, "*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*".

Linnet Ivet Bello Leyva en su tesis de pre-grado define la vigilancia tecnológica como: "*La actividad sistemática que permite identificar las evoluciones y novedades del entorno tecnológico, que facilite a las organizaciones orientar su actividad estratégica hacia el desarrollo de la innovación y anticipación de los cambios producidos en dicho entorno.*" [15]

A partir del análisis de los conceptos anteriormente expuestos se asume la definición brindada por Linnet Bello Leyva ya que constituye la base del proceso de VT definido para el centro FORTES, siendo indispensable su implementación e implantación en instituciones de I+D+i debido a los beneficios que tributan al éxito de la organización.

1.2.2 Beneficios de la VT

Aunque existen instituciones que no han sido conscientes en ocasiones de las ventajas de la vigilancia tecnológica, debido a que sus esfuerzos de captación y empleo de la información no se realizan sistemáticamente; la vigilancia está presente en un número cada vez mayor de estas. Su aplicación puede beneficiar al conjunto del ciclo innovador. A continuación se mencionan los beneficios consultados en la bibliografía por algunos autores.

Capítulo 1. Fundamentación Teórica

Según Ashton, Stacey, Palop y Vicente [9] plantean:

- Alerta sobre amenazas con repercusión en nuestro mercado desde sectores distintos al de la empresa.
- Ayuda a decidir el programa de I+D y su estrategia.
- Contribuye a abandonar a tiempo un determinado proyecto de I+D.
- Detecta oportunidades de inversión y comercialización. Su interrupción puede originar pérdida de mercados.
- Facilita la incorporación de nuevos avances tecnológicos a los propios productos y procesos.
- Identifica socios adecuados en proyectos conjuntos de I+D ahorrando inversiones.
- Permite evitar barreras no arancelarias ³en mercados exteriores.

En tanto Palop y Vicente, destacan [16]:

- **Anticiparse**, pues alerta sobre cambios o amenazas provenientes de sectores distintos al de la organización, le señala posibles nuevos nichos de mercado.
- **Reducir riesgos** al detectar competidores o productos entrantes o sustitutivos.
- **Ayuda en el proceso** de innovación porque ayuda a decidir el programa de I+D+i y su estrategia, igualmente a justificar el abandonar proyectos, resolver problemas específicos de carácter técnico.
- **Cooperación** al conocer nuevos socios, al establecer su capacidad e idoneidad de trabajo, también le permite identificar enlaces academia–sector productivo, entre otros.

Aunque la VT se practica usualmente en las empresas, también puede ser practicada en otras organizaciones, por ejemplo: instituciones educativas, centros tecnológicos, grupos de investigación, entidades estatales interesados en la investigación, desarrollo e innovación de procesos y productos.

³ Barreras técnicas a la distribución de productos por regulaciones de un país.

Capítulo 1. Fundamentación Teórica

En estas organizaciones la práctica de la VT trae beneficios como: [16]

- Puede ayudarles para alertarles sobre cambios en el desarrollo científico, señalando sectores emergentes de investigación.
- Les facilita la actualización de conocimientos, señala posibles nichos en los que puede llegar a competir el país u organización.
- Evita destinar recursos humanos y económicos en áreas cuya obsolescencia es inevitable.
- Puede ayudar a decidir el rumbo de las políticas de Ciencia Tecnología e Innovación, al igual que líneas de investigación.
- Permite identificar las capacidades de instituciones, grupos de investigadores e incluso empresas para llevar a cabo acuerdos de I+D+i.
- Permite identificar actores, como expertos de talla internacional o instituciones, que les sirvan como pares evaluadores o como posibles socios.

Por lo anteriormente expresado, en las instituciones la VT es una disciplina necesaria que les permitirá estar pendientes de los nuevos productos y actividades de los competidores, tanto internos como externos. Todo ello permite disponer de los beneficios mencionados y convertirse en una herramienta de toma de decisiones, en pro del desarrollo económico y social.

Luego de analizar los beneficios de la VT según los autores citados, se previenen entre las principales ventajas para el centro FORTES:

- Ayuda a decidir el programa de un proyecto de I+D y su estrategia.
- Anticipación y alerta sobre cambios o amenazas provenientes de sectores distintos al de la organización, señalando posibles nuevos nichos de mercado.
- Contribuye a abandonar a tiempo determinado proyecto de I+D.
- Reduce los riesgos al detectar competidores o productos entrantes o sustitutivos.
- Facilita la incorporación de nuevos avances tecnológicos a los propios productos y procesos.
- Identifica socios adecuados en proyectos conjuntos de I+D ahorrando inversiones.

Capítulo 1. Fundamentación Teórica

Para realizar la vigilancia es necesario consultar fuentes especializadas brindando la posibilidad de poseer información confiable y actualizada, de la cual la institución pueda nutrirse.

1.2.3 Fuentes de información

“Borghoff y Pareshi, distinguen el conocimiento + ‘tácito’ del ‘explícito’. El conocimiento **explícito** es el conocimiento **formal** que puede ser empaquetado como información. Este puede ser encontrado en los documentos de una organización: informes, artículos, manuales, imágenes, etc. El conocimiento **tácito** es el conocimiento personal que está arraigado en la experiencia individual y es compartido e intercambiado a través del contacto directo. El conocimiento tácito es la clave para tomar decisiones y desarrollar acciones.” [17]

Luego del análisis anterior las fuentes de información se pueden clasificar en formales e informales. Estas juegan un papel fundamental en el éxito del proceso de la vigilancia. Las bases de datos de patentes constituyen una magnífica fuente de información para practicar labores de vigilancia tecnológica. Aunque no se pueden descartar otras fuentes como son:

Fuentes formales:

- Patentes
- Publicaciones científicas y tecnológicas
- Estudios de mercado
- Informes técnicos y estadísticos de organismos oficiales
- Estándares
- Tesis

Fuentes informales:

- Prensa científico-técnica
- Noticias
- Ferias
- Clientes y proveedores

Con el desarrollo de la informática el uso de herramientas para las labores de vigilancia se ha hecho indispensable debido a la gran cantidad de información que se publica a diario, resultando difícil recuperarla de manera manual; el uso de la inteligencia artificial en las mismas permite ayudar a los investigadores en la toma de decisiones, representando un apoyo importante en el proceso.

Capítulo 1. Fundamentación Teórica

1.3 Herramientas para la VT

En la medida que la actividad de vigilancia e inteligencia incrementen su importancia para la organización, serán necesarias herramientas especializadas en las diferentes etapas del proceso de vigilancia. La organización deberá evaluar su necesidad e impacto en sus objetivos generales, y específicamente en los estratégicos, siendo importante conocer estas herramientas y estudiar la factibilidad de su uso o integración.

1.3.1 Principales tendencias de las herramientas de VT

“Las tendencias muestran que todas las etapas del proceso de vigilancia estarán influenciadas por las tecnologías multimedia. La etapa de análisis, como la más importante y compleja, recibe las mayores contribuciones: inteligencia en tiempo real gracias a la creación de interfaces interactivas, el desarrollo de software más potente de tratamiento de la información, y el desarrollo de técnicas que integran la presentación y la modelación, entre otras. En la entrega de la información resultante las direcciones se observarán en el uso de novedosas técnicas de visualización, de sistemas de comunicación electrónicos y vídeo-conferencias con encuentros cara a cara.” [18].

A continuación se ilustran las tendencias de las herramientas en cada una de las etapas del proceso de vigilancia tecnológica (Ver figura 5).

Principales tendencias en las herramientas de vigilancia e inteligencia

BÚSQUEDA Y RECUPERACIÓN	TRATAMIENTO Y ANÁLISIS	VISUALIZACIÓN, ENTREGA, VALORIZACIÓN
Uso creciente de las tecnologías multimedia (RSS, feed de alertas...)	Técnicas de presentación y modelación integradas.	Desarrollo en técnicas de visualización de los resultados: mapas tecnológicos, grafos relacionales, tablas, matrices...
Herramientas que permiten mejor acceso a fuentes de información de calidad: Internet Invisible, bases de datos de patentes, publicaciones, proyectos, mercados...	Softwares de procesamiento de más potentes: altos volúmenes, uso de técnicas estadísticas, minería de datos y de textos	Desarrollo de técnicas de comunicaciones electrónicas, videos conferencias y encuentros cara a cara.
Búsqueda en lenguaje natural, sistemas Question-Answering...	Técnicas que apoyan la gestión documental de las organizaciones: categorización y clasificación (clustering) de la información	Desarrollo de plataformas integrales de vigilancia e inteligencia, orientadas no solo a la recuperación y análisis de la información, sino a la entrega de resultados y a su valorización a partir del criterio y evaluación por parte de los usuarios
Acceso y recuperación de información en diversos formatos multimedia (texto, video, imagen, audio...), idiomas, perfiles...	Combinación Cienciometría/Bibliometría/Criterio expertos.	
Mejoramiento en las formas de almacenamiento de la información	Cómputo / interfase interactiva para visualización e inteligencia «just in time» como apoyo a la previsión tecnológica/mercado.	

Figura 5. Principales tendencias en las herramientas de vigilancia e inteligencia.

Fuente: Pilar Lázaro. La inteligencia competitiva, factor clave para la toma de decisiones estratégicas en las organizaciones. 2007-2008, Madrid. Plan Regional de Ciencia y Tecnología de la Comunidad de Madrid

Capítulo 1. Fundamentación Teórica

Existen herramientas de captura y tratamiento de la información que han potenciado el desarrollo de la VT haciendo uso de técnicas de la minería de datos y texto.

Según Dousset la Minería de Datos (Data Mining) es: *“El tratamiento automático de la información bruta contenida en las bases de datos que permite extraer los esquemas y los modelos más significativos con objeto de presentar a los usuarios conocimientos implícitos, no triviales, desconocidos anteriormente y potencialmente útiles”*. [18]

“La Minería de Textos (Text Mining) es el proceso de aplicación de métodos automáticos para analizar y estructurar datos de texto con el objetivo de crear un conocimiento útil a partir de información estructurada y no estructurada.” [18]

1.3.2 Herramientas internacionales

Tetralogie

Este paquete informático profesional fue desarrollado en el Instituto de Recherche Informatique de Toulouse (IRIT), Francia bajo la dirección del investigador Bernard Dousset. Tetralogie proporciona un conjunto de herramientas para realizar análisis cuantitativos y estudios de Vigilancia tecnológica desde las etapas de adquisición hasta la de interpretación de los datos. Este software se clasifica dentro del grupo que utiliza métodos estadísticos y algunos modelos de la lingüística, al incorporar una etapa de análisis morfológico (equiparación⁴ de términos similares).

La herramienta parte de que la fase de obtención de datos ha sido realizada, que los datos están disponibles y que se investigará un tema. Esto supone que la fuente de información es conocida y que su descripción está disponible. Cuando los datos provienen de una base de datos no estructurada o cuando existe el interés de estudiar un campo a texto completo dentro de una base de datos estructurada (Titulo o Abstract de INSPEC) el paquete Tetralogie dispone de una fase de pre-tratamiento, con modelos estadísticos y modelos morfológicos. [18]

Es muy útil para el tratamiento de grandes volúmenes de información provenientes de fuentes bibliográficas, tales como bases de datos de patentes y de publicaciones científicas. Utiliza el análisis de coocurrencia de palabras⁵, y permite la visualización de la

⁴ Consideración de dos o más cosas o personas como equivalentes al compararlas (<http://es.thefreedictionary.com/equiparaci%C3%B3n>)

⁵ Unión de dos palabras que se relacionan o tienen alguna proximidad entre ellas.

Capítulo 1. Fundamentación Teórica

información en mapas vectoriales de gran precisión, donde la proximidad entre palabras adquiere gran significado.

Matheo Software

Son herramientas de minería de datos (recuentos y coocurrencia de palabras), desarrolladas por el grupo de origen francés Matheo Software.

El software **Matheo Analyzer** permite procesar colecciones documentales con información estructurada de una forma rápida. Analiza la coaparición de palabras y efectúa la visualización de los resultados a través de histogramas y grafos (formas, pares y conectividad).

Facilita la creación y filtrado de diccionarios de trabajo, la importación de cualquier índice o tesoro que el analista quiera utilizar. Visualiza los patrones y tendencias en grafos de relaciones simétricos y asimétricos.

El software **Matheo Patent** es una versión del Matheo Analyzer, orientada específicamente al tratamiento en línea de dos colecciones documentales de patentes: la colección de la oficina de los Estados Unidos (USPTO) y la colección de la Oficina Europea de Patentes (EPO).

Matheo Web ha sido diseñado para la búsqueda, recuperación y análisis de fuentes y recursos de información sobre Internet (web, mails, noticias...). El software permite la búsqueda mediante palabras claves, y facilita la descarga automática de las páginas web, y además actualiza las peticiones de búsqueda. Incluye soporte estadístico para el análisis de la información, y un motor de búsqueda local que permite el análisis y visualización de la información identificada. [18]

La comparación de estas herramientas se puede consultar en el Anexo 2.

Vigtech

La herramienta fue creada por la Universidad Nacional de Colombia; automatiza los procesos de captación y búsqueda de datos mediante el módulo *crawler-vigtech* que permite descargar los documentos científicos de SCOPUS, y extraer características de dichos documentos, construyendo así, una base de datos relacional en la cual se almacenan estructuradamente los metadatos⁶ del artículo y del autor.

⁶ Según Howe (1993), el término fue acuñado por Jack Myers en la década de los 60 para describir conjuntos de datos. La primera acepción que se le dio (y actualmente la más extendida) fue la de dato sobre dato.

Capítulo 1. Fundamentación Teórica

Esto permite realizar análisis descriptivos y análisis exploratorios de datos que apoyan la fase de análisis y organización. La herramienta informática utiliza técnicas de aprendizaje de máquina y de minería de datos apoyando así la fase de inteligencia, utilizando algoritmos para análisis de redes sociales, reducción de dimensionalidad, escalamiento multidimensional, agrupamiento, modelos gráficos probabilísticos, entre otros, que permiten vincular de una forma inteligible los resultados obtenidos presentando indicadores, mapas, sociogramas y en general representaciones relacionales de un tópico dado.

El desarrollo de la herramienta informática está basado en una licencia GPL. Se utilizó la arquitectura cliente servidor. Como servidor web se utilizó Apache 2.0 Handler, y como gestor de base de datos MySQL 4.0.18. Los desarrolladores utilizaron PHP versión 4.3.4 y Java Script. [19]

Goldfire

“Se trata de una herramienta creada por Invention Machine, una compañía de Estados Unidos de desarrollo de software. Es una herramienta mixta que incluye tanto componentes de minería de datos para el análisis de la actividad tecnológica (principalmente patentes), como también módulos para el tratamiento textual, tales como diccionarios de conceptos científicos que puede incorporar. Además incluye varias herramientas y métodos para el soporte en la resolución de problemas tales como TRIZ.” [18]

1.3.3 Herramientas nacionales

DataSOMining (Instituto Finlay, Universidad Nacional Autónoma de México)

“Software de minería de datos con un enfoque neurocomputacional que permite entre sus principales funciones: la adquisición de datos contenidos en la base de datos Medline; la normalización de distintos campos de forma automática, mediante el uso de tesauros o manual; creación de distintas matrices numéricas que servirán como datos de entrada para la red neuronal; el entrenamiento de la red a partir de una determinada configuración y la generación de visualizaciones mediante mapas de clustering de los datos entrenados mediante la red SOM.” [20]

<http://www.scielo.br/pdf/ci/v32n2/17038.pdf>

Capítulo 1. Fundamentación Teórica

BioMundiPatent (Consultoría Biomundi/IDICT)

“Herramienta desarrollada para búsqueda y la descarga de patentes provenientes de la base de datos de la Oficina Española de Patentes y Marcas (Espacenet) con un formato compatible para ser importado en un sistema de gestión de base de datos bibliográfica como el Procite, EndNote o Reference Manager. Además brinda la posibilidad de obtener los resultados con los campos útiles para el análisis en una base de datos en Microsoft Office Excel. Permite la homogeneización del campo Código Internacional de Patentes (CIP), eliminando los duplicados.” [20]

Prolntec (Universidad de Pinar del Río)

“Sistema modular *ad hoc* que integra las etapas para el análisis de la información proveniente de bases de datos de patentes. Se desarrolló con lenguaje script PHP, PostgreSQL como gestor de base de datos y Apache como servidor web. Además utiliza de forma transparente para el usuario el programa NetDraw 2.062 para mostrar las representaciones visuales. Permite la búsqueda, recuperación, procesamiento, análisis y presentación de información contenida en documentos de patentes procedente de bases de datos internacionales.” [20]

CiteLivePro 2.0 (Centro de Bioplantas, Universidad de Ciego de Ávila)

“Programa diseñado para brindar servicio en ambiente Web para la organización bibliográfica, y entre sus opciones se encuentran: la búsqueda bibliográfica en bases de datos, que pueden ser propietarias, o descargas de bases de datos en formato .txt o convertibles a esta extensión; creación de bases de datos personalizadas; generación automática de bibliografías acorde a cualquiera de las normas bibliográficas que posea el sistema; descargas de documentos a texto completo mientras lo permitan las regulaciones de la propiedad intelectual; por último permite la búsqueda por cualquiera de los términos que contiene la bases de datos.” [20]

InfoCam v3.0

“Es el resultado de investigaciones y aplicaciones de un equipo multidisciplinario de varios profesionales en informática, bibliotecología, ciencia de la información y las ciencias matemáticas. Se caracteriza por ser un software de propósito especial, diseñado para aplicaciones de métodos cuantitativos y de modelos bibliométricos, informétricos y cientiométricos en las tareas de análisis de información. Sus capacidades y características, van desde la normalización de bases de datos; la edición de glosarios

Capítulo 1. Fundamentación Teórica

terminológicos estructurados según relaciones paradigmáticas; la conversión de referencias bibliográficas descargadas de los proveedores en línea; gama de gráficos para la visualización, la instrumentación del algoritmo de Kohonen para la elaboración de mapas; la creación de nuevos conjuntos de registros específicos, a partir de los resultados informétricos, o deseos expresos del investigador, para el estudio más específico—detallado—de conjuntos de científicos, instituciones, países y/o documentos. Ofrece la posibilidad de imprimir salidas en forma de tablas matriciales y gráficas, diseñadas por el usuario; entre otras aplicaciones.” [20]

Mimosa (OCPI)

“Combina los procesos de descarga automatizada de registros de bases de datos bibliográficas con el procesamiento de los mismos, hasta lograr en una interface adecuada para el posterior trabajo de procesamiento analítico de la misma. Mediante el uso de los módulos correspondientes a este sistema, se realiza la estructuración y homogenización de los formatos de salida de la información de cada base de datos, y permite obtener archivos con la estructura apropiada para su inclusión en los llamados sistemas de gestión de bases de datos como el Procite, el Reference Manager o el EndNote, o de lo contrario directamente a los sistemas de procesamiento matemático y estadístico de la información.” [20]

Sistema automatizado para la vigilancia de patentes SiVigPat (CENIC)

“Sistema que permite procesar la información extraída de diversas bases de datos de libre acceso de patentes en Internet. Entre sus opciones se encuentran: importar y editar patentes con la asignación de descriptores; identificación de patentes análogas o equivalentes, así como de las familias de patentes; clasificación de patentes en solicitadas y solicitadas-concedidas; Importación y análisis de citas; conteo automatizado de: descriptores asignados, titulares, autores, códigos de clasificación; países de prioridad, de derecho y designados; origen y destino de la tecnología. Representación en forma de tablas o en gráficos con valores absolutos o acumulativos.” [20]

Toollnf (Consultoría Biomundi/IDICT)

“Herramienta para el análisis y procesamiento de la información bibliográfica, mediante opciones de conteo, identificación de datos, clasificación de registros y creación y normalización de matrices de coocurrencia. Su ventaja principal se enfoca en la realización de conteos automatizados, que pudieran hacerse mediante las tablas

Capítulo 1. Fundamentación Teórica

dinámicas de Excel, con la particularidad de que se cuentan por separado elementos que se encuentran dentro de una misma celda y se realizan conteos clasificatorios, en los cuales no sólo se obtiene la cantidad de registros en que aparece cada elemento, sino también el número de los registros. Además, es muy útil en la realización de matrices de coocurrencia para campos con más de un elemento por celda, ya que se pueden combinar dos hojas de conteos clasificatorios en una matriz. Es una herramienta que se instala como un complemento a Microsoft Excel, por lo que su utilización es sencilla y eficaz.” [20]

Centro de Información y Gestión Tecnológica (CIGET), Villa Clara

“El software permite la descarga de los documentos de patentes localizados en la base Espacenet y las importa a una base de datos en Access, donde se realiza el procesamiento primario, hasta llegar a la homogeneización de los datos, para extraer principales países generadores de patentes en la temática, países designados, registros por años, principales inventores, principales firmas, temáticas más destacadas dentro de la tecnología y su evolución en el tiempo. Los gráficos de salida se pueden editar pero solo estéticamente, no conceptualmente. La herramienta permite además relacionar campos, por ejemplo: inventor-país y compañía-país.” [20]

Valoración sobre las herramientas

A pesar de que todas las herramientas anteriormente analizadas incluyen los aspectos esenciales en el proceso de VT, no constituyen la solución para las necesidades del grupo GIC; las estudiadas a nivel nacional, para el procesamiento de la información tienen en cuenta algunas de las funcionalidades referentes al tema como son análisis bibliométricos, cuantitativos, entre otros.; pero, no existe una herramienta que contemple la mayoría de los tipos de análisis que hoy existen.

Las examinadas a nivel internacional contemplan el análisis y procesamiento de la información de fuentes específicas, pero la mayoría son propietarias, costosas y de uso personal.

Vigtech es una herramienta gratuita de las más completas en el análisis de la información pero tiene como restricción el uso exclusivo de SCOPUS, una de las bases de datos de resúmenes más grande del mundo, que tiene la limitante que es de pago por lo que no constituye nuestra solución.

Capítulo 1. Fundamentación Teórica

El estudio de las herramientas responde a funcionalidades que se podrán retomar en la elaboración del diseño del sistema de vigilancia, logrando de esta manera que el mismo se encuentre a la altura de los sistemas de su tipo a nivel nacional e internacional utilizando términos modernos como la bibliometría, cienciometría, análisis de redes sociales y minería de datos.

1.4 Bibliometría y cienciometría

En las actividades de vigilancia se utilizan de manera creciente las aportaciones de la bibliometría y la cienciometría para realizar el tratamiento de grandes volúmenes de información con la ayuda de herramientas informáticas. A menudo, ambos términos se utilizan indistintamente, aunque sus diferencias son claras. La bibliometría tiene por objeto estudiar los libros y revistas científicas así como comprender las actividades de comunicación de la información. En cambio, la cienciometría se centra en el estudio de aspectos cuantitativos de la creación, difusión y utilización de la información científica con el objetivo de comprender los mecanismos de investigación como actividad social. Entre otros indicadores, se utilizan el recuento de publicaciones y las citas entre artículos.

“La cienciometría parte de la base de que los resultados de las investigaciones científicas y técnicas se plasman en forma escrita a través de artículos de revistas, memorias de patentes, actas de congreso, entre otros. El progreso científico y técnico queda en escritos. El análisis de esta documentación escrita permitirá, descubrir sus características y su evolución”. [12]

“La cienciometría se basa en el análisis y cómputo de determinados indicadores bibliométricos como: autores de artículos, citas que aparecen en la bibliografía de cada artículo, palabras (keywords) contenidas en los títulos de los artículos o en los resúmenes. Estos están presentes en los registros de las bases de datos” [12]. Mediante el recuento de estos indicadores se puede determinar, por ejemplo:

- El crecimiento de cualquier campo de la ciencia, según la variación cronológica del número de trabajos publicados en él.
- El envejecimiento de los campos científicos, según la vida media de las referencias de sus publicaciones.
- La evolución cronológica de la producción científica, según el año de la publicación de los documentos.
- La productividad de los autores o instituciones, medida por el número de sus trabajos.

Capítulo 1. Fundamentación Teórica

- La colaboración entre los científicos o instituciones, medida por el número de autores por trabajo o centros de investigación que colabora.
- El impacto o visibilidad de las publicaciones dentro de la comunidad científica internacional, medido por el número de citas que reciben estas por parte de trabajos posteriores.
- El análisis y evaluación de las fuentes difusoras de los trabajos, por medio de indicadores de impacto de las fuentes.
- Las trayectorias tecnológicas seguidas por empresas o países en un período determinado, de acuerdo con su actividad patentadora.

1.5 Tecnologías para el diseño

En este epígrafe se analizan y seleccionan las tecnologías para el diseño del sistema, como son: lenguaje de modelado, herramientas de Ingeniería de Software Asistida por Computación (CASE) y metodologías de desarrollo de software. Asegurando con su elección la calidad del diseño.

1.5.1 Lenguaje de modelado UML 2.1

El Lenguaje Unificado de Modelado (UML), por sus siglas en inglés (Unified Modeling Language), se define como un lenguaje que permite especificar, visualizar y construir los artefactos de los sistemas de software; constituye un proceso sólido para la construcción de modelos. UML es un estándar, su utilización es independiente del lenguaje de programación y de las características de los proyectos, pues ha sido diseñado para modelar cualquier tipo de proyecto, tanto informáticos como de arquitectura, o de cualquier otra rama. También intenta solucionar el problema de propiedad de código que se da con los desarrolladores, al implementar un lenguaje de modelado para todos los desarrollos, se crea una documentación común, que cualquier desarrollador con conocimientos de UML será capaz de entender, independientemente del lenguaje de programación utilizado para el desarrollo.

1.5.2 Herramientas de Ingeniería de Software Asistida por Computadora (CASE)

CASE corresponde a las iniciales de: Computer Aided Software Engineering. De acuerdo con Kendall y Kendall: *“La ingeniería de sistemas asistida por ordenador es la aplicación de tecnología informática a las actividades, las técnicas y las metodologías propias de desarrollo, su objetivo es acelerar el proceso para el que han sido diseñadas, en el caso de CASE para automatizar o apoyar una o más fases del ciclo de vida del desarrollo de sistemas”*. [21]

Capítulo 1. Fundamentación Teórica

Estas herramientas permiten a los desarrolladores modelar y documentar sus artefactos, cubriendo el ciclo de vida del proceso de desarrollo de software. Una herramienta CASE puede incluir: diccionario de datos, herramientas de diseño, herramientas de desarrollo de modelo de datos y herramientas para el desarrollo de prototipos. Para el desarrollo de la solución informática es necesario utilizar algunas herramientas para facilitar el trabajo y por consiguiente realizarlo con la calidad requerida.

1.5.2.1 Visual Paradigm for UML 8.0

Visual Paradigm es una herramienta profesional que soporta el ciclo de vida completo del desarrollo de software. El software de modelado ayuda a una rápida construcción de aplicaciones de calidad, mejores y a un menor coste. Permite construir todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación.

Ofrece [22]:

- Entorno de creación de diagramas para UML 2.1.
- Diseño centrado en casos de uso y enfocado al negocio que genera un software de mayor calidad.
- Uso de un lenguaje estándar común a todo el equipo de desarrollo que facilita la comunicación.
- Capacidades de ingeniería directa (versión profesional) e inversa.
- Modelo y código que permanece sincronizado en todo el ciclo de desarrollo.
- Disponibilidad de múltiples versiones, para cada necesidad.
- Disponibilidad de integrarse con los principales entorno de desarrollo integrado (IDEs).
- Disponibilidad en múltiples plataformas.

1.5.2.2 Rational Rose Enterprise 7.0

Es una herramienta desarrollada por los creadores de UML que cubre todo el ciclo de vida de un proyecto, proporcionando un lenguaje común de modelado que facilita la creación de software con calidad.

Características adicionales incluidas. [23]

Capítulo 1. Fundamentación Teórica

- Soporte para análisis de patrones ANSI C++, Rose J y Visual C++ basado en "Design Patterns: Elements of Reusable Object-Oriented Software".
- Característica de control por separado de componentes modelo que permite una administración más granular y el uso de modelos.
- Soporte de ingeniería Forward y/o reversa para algunos de los conceptos más comunes de Java 1.5.
- La generación de código Ada, ANSI C ++, C++, CORBA, Java y Visual Basic, con capacidad de sincronización modelo- código configurables.
- Soporte Enterprise Java Beans 2.0.
- Capacidad de análisis de calidad de código.
- El Add-In para modelado Web provee visualización, modelado y las herramientas para desarrollar aplicaciones web.
- **Modelado UML** para trabajar en diseños de base de datos, con capacidad de representar la integración de los datos y los requerimientos de aplicación a través de diseños lógicos y físicos.
- Capacidad de crear definiciones de tipo de documento XML (DTD) para el uso en la aplicación.
- Integración con otras herramientas de desarrollo de Rational.
- Capacidad para integrarse con cualquier sistema de control de versiones SCC-compliant, incluyendo a Rational ClearCase.
- Publicación web y generación de informes para optimizar la comunicación dentro del equipo.

Consideraciones de las Herramientas CASE

A partir del estudio realizado sobre las principales herramientas CASE analizadas, se decidió utilizar Visual Paradigm for UML v8.0. Se hace esta elección, principalmente, por ser una herramienta multiplataforma, permite guardar todo el modelo en un solo fichero facilitando tener todo el trabajo encapsulado, es una herramienta que se desempeña bien sobre ordenadores "de bajas prestaciones.", permite exportar documentos, es robusta y de fácil uso. Además, se sustenta la elección a partir de que la UCI cuenta con la licencia para su uso.

Capítulo 1. Fundamentación Teórica

1.5.3 Metodologías de desarrollo de software

“La metodología de desarrollo de software en ingeniería de software es un marco de trabajo usado para estructurar, planificar y controlar el proceso de desarrollo en sistemas de información calidad”. [24]. En un proyecto de software la metodología de desarrollo define ¿Quién debe hacer qué?, y ¿Cuándo debe realizarlo?

Una metodología está compuesta en:

- Cómo dividir un proyecto en etapas.
- Qué tareas se llevan a cabo en cada etapa.
- Qué restricciones deben aplicarse.
- Qué técnicas y herramientas se emplean.
- Cómo se controla y gestiona un proyecto.

1.5.3.1 Rational Unified Process (RUP) [25]

RUP es una forma disciplinada de asignar tareas y responsabilidades en una empresa de desarrollo de software que define quién hace qué, cuándo y cómo.

Costo del cambio:

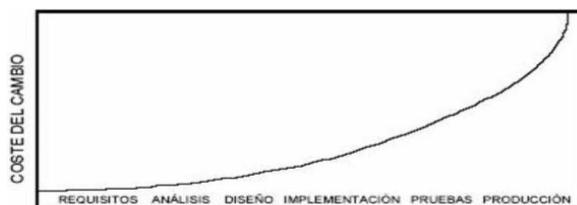


Figura 6. RUP Costo del cambio.

Fuente: Universidad de San Martín de Porres.

Comparando RUP Y XP.2012. <http://www.usmp.edu.pe/>

Como se puede ver en la figura anterior un cambio en las etapas de vida del sistema incrementaría notablemente el costo.

Requiere un grupo grande de programadores para trabajar con esta metodología.

Es un marco del proyecto que describe una clase de los procesos que son iterativos e incrementales. Define las actividades y los artefactos que se necesitan. Es el proceso de desarrollo más general de los existentes actualmente.

Capítulo 1. Fundamentación Teórica

Los **procesos** estiman tareas y horario del plan midiendo la velocidad de iteraciones concerniente a sus estimaciones originales. Las iteraciones tempranas de proyectos conducidos por RUP se enfocan fuertemente sobre la arquitectura del software, la puesta en práctica rápida de características se retrasa hasta que se ha identificado y se ha probado una arquitectura firme, presta énfasis en los requisitos y el diseño.

La ventaja principal de la metodología es que se basa en las mejores prácticas que se han intentado y se han probado en el campo.

RUP se divide en cuatro fases (Ver figura 7):

- Inicio (define el alcance del proyecto).
- Elaboración (definición, análisis, diseño).
- Construcción (implementación).
- Transición (fin del proyecto y puesta en producción).

Cada fase concluye con un **hito** (toma de decisiones).

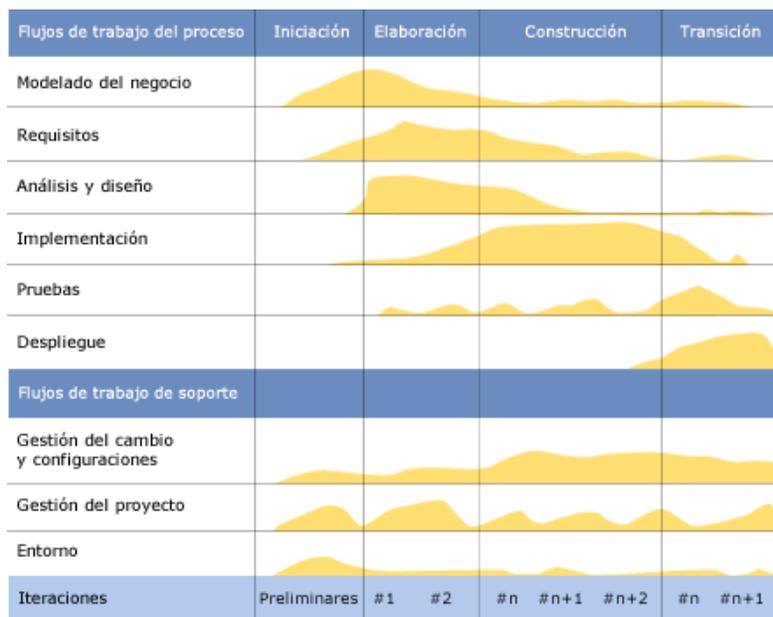


Figura 7. Flujos de trabajo de RUP.

Fuente: Universidad Bicentenario de Aragua

Metodología RUP.2012. <http://www.usmp.edu.pe/>

Capítulo 1. Fundamentación Teórica

Las características principales de RUP son:

Guiado / Manejado por casos de uso: La razón de ser de un sistema software es servir a usuarios ya sean humanos u otros sistemas; un caso de uso es una facilidad que el software debe proveer a sus usuarios.

Centrado en arquitectura: La arquitectura involucra los elementos más significativos del sistema y está influenciada entre otros por plataformas de software, sistemas operativos, manejadores de bases de datos, protocolos, consideraciones de desarrollo como sistemas heredados y requerimientos no funcionales.

Iterativo e Incremental: Para hacer más manejable un proyecto se recomienda dividirlo en ciclos. Para cada ciclo se establecen fases de referencia, cada una de las cuales debe ser considerada como un mini proyecto cuyo núcleo fundamental está constituido por una o más iteraciones de las actividades principales básicas de cualquier proceso de desarrollo. En concreto RUP divide el proceso en cuatro fases, dentro de las cuales se realizan varias iteraciones en número variable según el proyecto y en las que se hace un mayor o menor hincapié en las distintas actividades.

Entre otras se pueden mencionar:

- RUP realiza un levantamiento exhaustivo de requerimientos.
- Busca detectar defectos en las fases iniciales.
- Intenta reducir al número de cambios tanto como sea posible.
- Realiza el Análisis y diseño, tan completo como sea posible.
- Diseño genérico, intenta anticiparse a futuras necesidades.
- Las necesidades de clientes no son fáciles de discernir.
- Existe un contrato prefijado con los clientes.
- El cliente interactúa con el equipo de desarrollo mediante reuniones.

1.5.3.2 Extreme Programming (XP) [25]

La programación extrema o eXtreme Programming (XP) es un enfoque de la ingeniería de software formulado por Kent Beck y De Jean, Extreme Programming Explained. Es la más destacada de los procesos ágiles de desarrollo de software. Se diferencia de las metodologías tradicionales principalmente en que pone más énfasis en la adaptabilidad que en la previsibilidad.

Capítulo 1. Fundamentación Teórica

XP surge en busca de simplificar el desarrollo del software y que se logre reducir el costo del proyecto.

Costo de cambio:

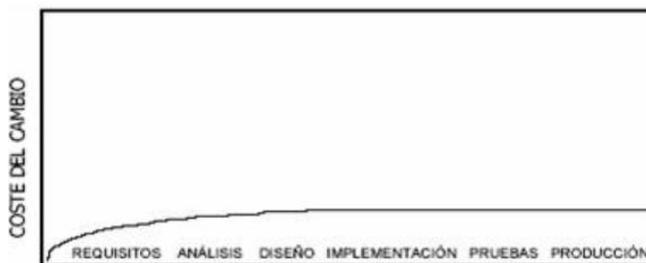


Figura 8. XP Costo del cambio.

Fuente: Universidad de San Martín de Porres.
Comparando RUP Y XP.2012. <http://www.usmp.edu.pe/>

Como se puede ver en la figura anterior un cambio en las etapas de vida del sistema reduce notablemente el costo.

Características de la metodología XP. [25]

- Desarrollo iterativo e incremental: pequeñas mejoras, unas tras otras.
- Pruebas unitarias continuas, frecuentemente repetidas y automatizadas, incluyendo pruebas de regresión. Se aconseja escribir el código de la prueba antes de la codificación.
- Programación en parejas: se recomienda que las tareas de desarrollo se lleven a cabo por dos personas en un mismo puesto. Se supone que la mayor calidad del código escrito de esta manera -el código es revisado y discutido mientras se escribe- es más importante que la posible pérdida de productividad inmediata.
- Frecuente integración del equipo de programación con el cliente o usuario. Se recomienda que un representante del cliente trabaje junto al equipo de desarrollo.
- Corrección de todos los errores antes de añadir nueva funcionalidad. Hacer entregas frecuentes.
- Refactorización del código: reescribir ciertas partes del código para aumentar su legibilidad y mantenibilidad pero sin modificar su comportamiento. Las pruebas han de garantizar que en la refactorización no se ha introducido ningún fallo.
- Propiedad del código compartida: en vez de dividir la responsabilidad en el desarrollo de cada módulo en grupos de trabajo distintos, este método promueve

Capítulo 1. Fundamentación Teórica

el que todo el personal pueda corregir y extender cualquier parte del proyecto. Las frecuentes pruebas de regresión garantizan que los posibles errores serán detectados.

- Simplicidad en el código: cuando todo funcione se podrá añadir funcionalidad si es necesario. La programación extrema apuesta que es más sencillo hacer algo simple y tener un poco de trabajo extra para cambiarlo si se requiere, que realizar algo complicado y quizás nunca utilizarlo. La simplicidad y la comunicación son extraordinariamente complementarias.

Es una metodología ágil centrada en potenciar las relaciones interpersonales como clave para el éxito en el desarrollo de software, promoviendo el trabajo en equipo, preocupándose por el aprendizaje de los desarrolladores, y propiciando un buen clima de trabajo. Se define especialmente adecuada para proyectos con requisitos imprecisos y muy cambiantes, y donde existe un alto riesgo técnico.

XP tiene una debilidad cuando se utiliza en dominios de aplicaciones complejas o situaciones difíciles en la organización: el rol del cliente no refleja los diferentes intereses, habilidades y fuerzas a las que enfrentan los programadores durante el desarrollo de proyectos.

Capítulo 1. Fundamentación Teórica

Partes de XP:

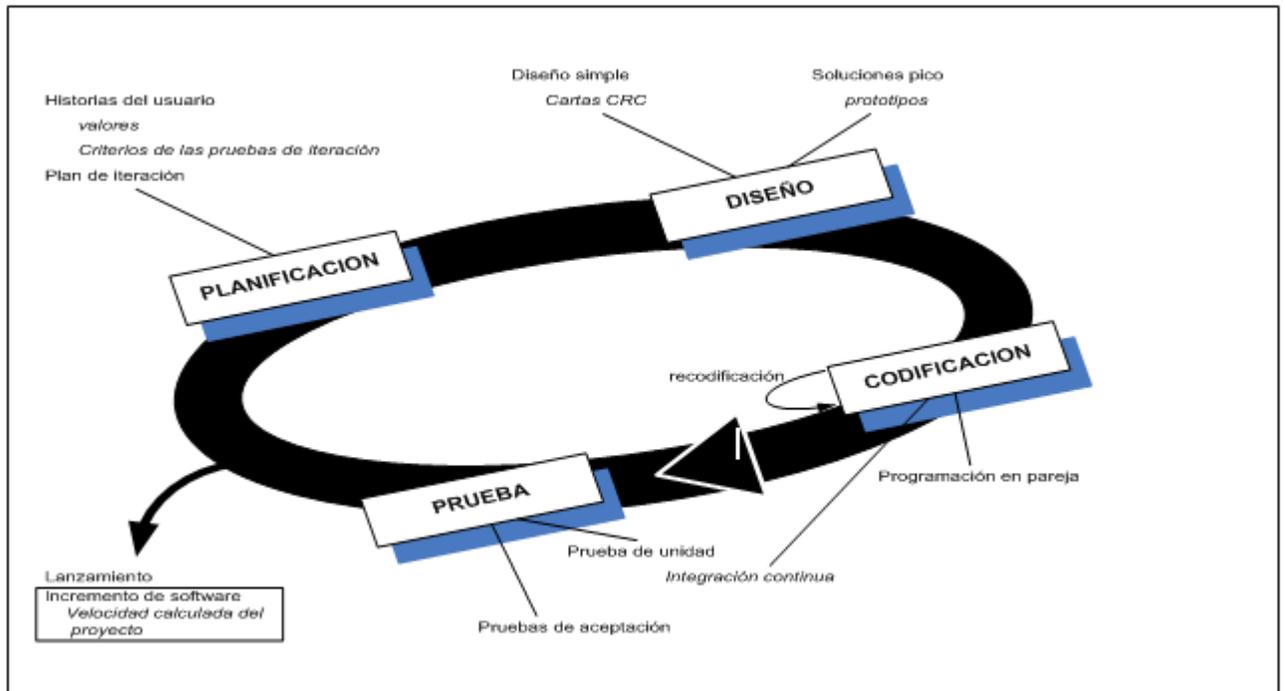


Figura 9. Partes de XP.

Fuente: Limas. Maestría en Ingeniería de Sistemas con Mención en Gestión En Tecnología de la Información, 2007

En la figura anterior se pueden observar las diferentes etapas que XP plantea para el desarrollo del software así como los artefactos que se generan.

A modo de resumen XP no se basa en la planificación, el cliente forma parte del equipo de desarrollo del software, existen constantes cambios en los requisitos y es una metodología ágil apropiada para proyectos pequeños.

Metodología seleccionada

Después de realizar el estudio de las principales metodologías de desarrollo de software se seleccionó RUP debido a que sus características son factibles para el desarrollo del sistema. Es recomendada para emprender proyectos complejos y de larga duración generando una serie de artefactos que sustentan el proceso de desarrollo de software. Es una metodología que se basa en la documentación por lo que no son deseados los cambios adaptándose al sistema a diseñar. Utiliza el lenguaje de modelado UML para la especificación, documentación y generación de los artefactos a obtener en el transcurso del desarrollo del proyecto. Documenta todo el proceso de análisis y diseño, lo cual es de mucha utilidad a la hora de implementar el sistema.

Capítulo 1. Fundamentación Teórica

Conclusiones parciales

- En la actualidad la automatización de procesos se apoya en gran medida en las Tecnologías de la Información y las Comunicaciones, utilizando sistemas informáticos como soporte para sus actividades, convirtiéndose en un elemento imprescindible de los sistemas de vigilancia tecnológica actuales.
- La minería de datos es muy utilizada en las herramientas de vigilancia tecnológica, debido a que permite la adecuada gestión de grandes volúmenes de información que se genera en el contexto de las organizaciones.
- La vigilancia tecnológica favorece el proceso de innovación ayudando a decidir el programa de I+D+i proporcionado líneas estratégicas de investigación e identificación de tendencias.

Capítulo 2. Propuesta del Sistema

2. CAPÍTULO 2. PROPUESTA DEL SISTEMA

En el presente capítulo se describen las etapas de vigilancia tecnológica que el sistema englobará, las funcionalidades a automatizar. Se presenta una propuesta del sistema con los requisitos funcionales y no funcionales a cumplir en la realización de la misma así como los roles del sistema y los artefactos generados por la metodología.

2.1 Descripción del sistema de vigilancia tecnológica

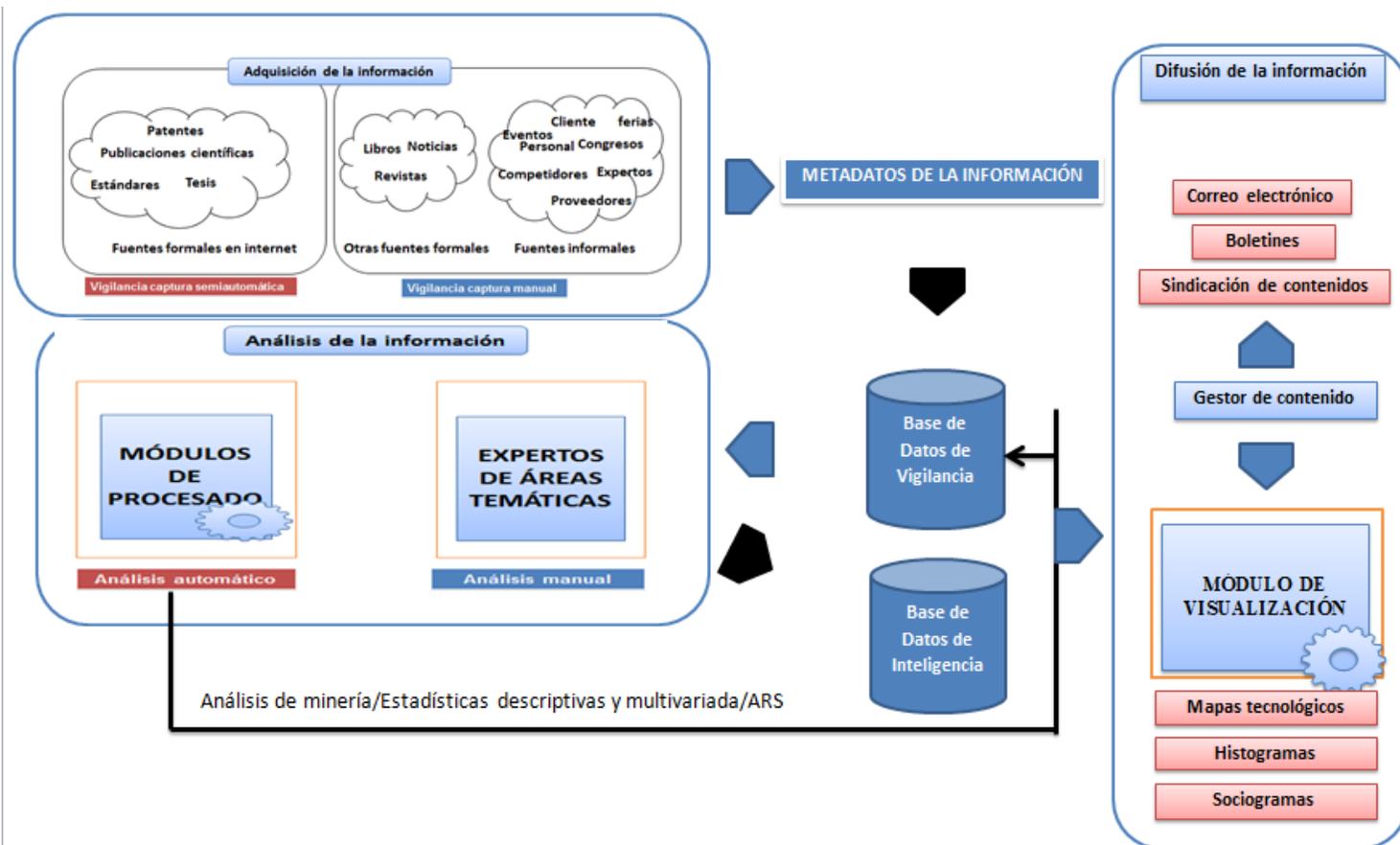


Figura 10. Esquema de sistema de vigilancia tecnológica para el grupo GIC.

Fuente: Elaboración propia.

Como se puede observar en la figura 10, el sistema de vigilancia realizará la adquisición de la información de fuentes formales e informales. En el caso de las fuentes formales el proceso será de forma semiautomática en la cual se extrae información de fuentes especializadas. Para los sitios web de revistas o publicaciones científicas será a través de redifusión web o sindicación de contenidos (RSS), en caso de bases de datos se aprovechará la posibilidad que estas brindan de exportar el resultado y los RSS; en las

Capítulo 2. Propuesta del Sistema

fuentes informales o no estructuradas así como las semi-estructuradas la extracción de la información se realizará de forma manual, en todos los casos se extraerán los metadatos y a través del sistema se guardarán para cada tipo de fuente los datos del autor, el artículo, la patente, entre otros.

El análisis de la información será concebido de forma manual y automatizada. Los expertos temáticos del Centro ejecutan un análisis de la información de las fuentes contenidas en la base de datos de vigilancia, arrojando informes de vigilancia tecnológica que serán guardados en la base de datos de inteligencia. El análisis de la información de forma automática se realizará a través de módulos de estadística descriptiva, análisis de redes sociales y minería de datos para la extracción de conocimiento, a partir de grandes volúmenes de documentos almacenados en la base de datos de vigilancia.

Para la difusión de la información se utilizará un gestor de contenidos el cual permitirá enviar alertas por correo electrónico según la necesidad de información a los usuarios, la sindicación de contenidos, publicación de boletines e informes de vigilancia tecnológica y además un módulo de visualización que permitirá graficar mapas tecnológicos, histogramas y sociogramas de relaciones entre autores o instituciones.

2.1.1 Adquisición de la información

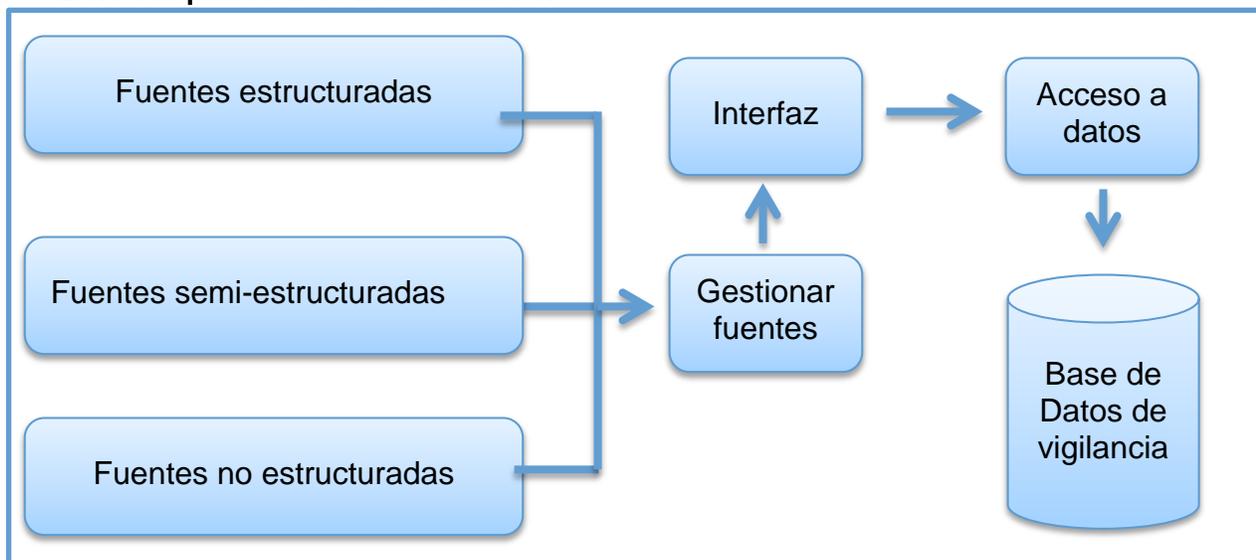


Figura 11. Proceso de adquisición de la información.

Fuente: Elaboración propia

El proceso de adquisición de la información se realizará según la figura anterior y que se explicará a continuación.

Capítulo 2. Propuesta del Sistema

“Las fuentes no estructuradas son aquellas en cuya construcción el investigador no ha tenido injerencia alguna y debe, por ende, realizar una faena de estructuración que le permita incorporarlas, como material, en su trabajo. Por el contrario, las fuentes estructuradas o semi- estructuradas son aquellas en las que el investigador intervino para adaptar las fuentes a sus necesidades de manera tal que la manifestación surgida de ellas se adapte bien a las exigencias de su tarea.” [26]

Gestionar fuentes: Permitirá de acuerdo a cada tipo de fuentes como: patente, tesis, estándares, publicaciones científicas, artículos de revista, noticias, libros, cliente, ferias, entre otros, guardar, modificar o eliminar los metadatos disponibles en la base de datos de vigilancia. Estos se seleccionarán según la relevancia que representen los mismos para el análisis y procesamiento de la información.

Se hará uso de una interfaz de usuario donde se debe seleccionar el tipo de fuente; el sistema será capaz de mostrar los campos correspondientes al mismo, que permitirá insertar estas fuentes de información haciendo uso de una clase de acceso a datos para el intercambio con la base de datos relacional.

2.1.2 Análisis y procesamiento de la información

El análisis de la información de forma manual se realizará por los expertos de las áreas temáticas del grupo GIC. Estos accederán a la información recopilada en la base de datos de vigilancia y formularán informes de vigilancia tecnológica para las diferentes áreas en dependencia de las necesidades, que luego serán almacenados en la base de datos de inteligencia. (Ver figura 12)

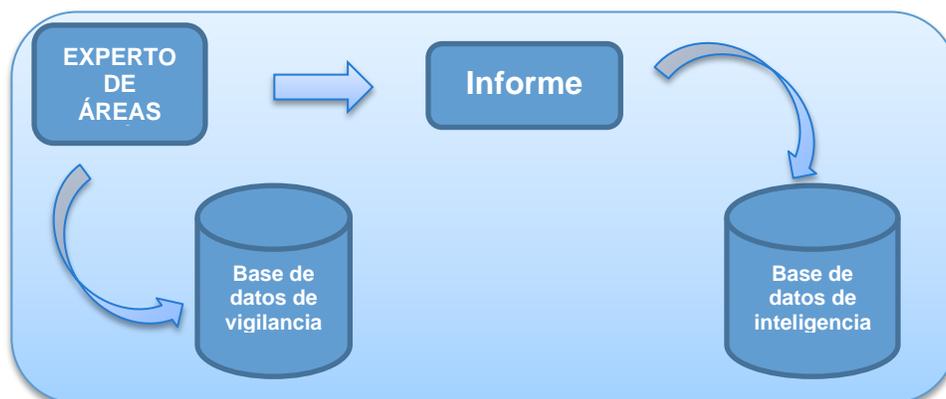


Figura 12. Proceso de análisis de la información manual.

Fuente: Elaboración propia

Capítulo 2. Propuesta del Sistema

Por otra parte el sistema debe procesar los datos siendo capaz de:

- Extraer características e información útil de los documentos
- Análisis de estadísticas descriptivas y multivariadas
- Análisis de redes sociales (ARS)



Figura 13. Proceso de análisis de la información automatizada.

Fuente: Elaboración propia

El procesamiento y análisis de la información se formalizará como ilustra la figura anterior que se explica a continuación.

2.1.2.1 Extracción de características e información útil de los documentos

El sistema debe extraer características de los documentos científicos tales como: palabras claves, temáticas y autores. De esta forma identificará agrupamientos naturales de los documentos por estos indicadores. Encontrará coocurrencias de palabras en los documentos en los campos títulos, resumen, descripciones para así analizar la proximidad entre ellas ya que pueden definir un área de estudio en la que ambas intervienen.

2.1.2.2 Análisis de estadísticas descriptivas y multivariadas

Se utilizan para describir el comportamiento de un conjunto de datos, siendo capaz de obtener indicadores como número de investigadores, número de artículos por año, número de artículos por autor, autores de una temática determinada, o indicadores bibliográficos sobre el estado de la publicación científica, palabras claves, estudio de referencias utilizando variables como autores, palabras clave, revista donde fue publicado, fechas, palabras encontradas en el artículo o patente, entre otras.

Para realizar los recuentos de indicadores bibliométricos como palabras clave es necesario lematizar las mismas. *“Los algoritmos de lematización realizan un proceso de normalización lingüística donde las diferentes formas que puede adoptar una palabra son reducidas a una*

Capítulo 2. Propuesta del Sistema

única forma común, a la cual se denomina stem o lema” [27]. Esto permitirá contar las palabras que conceptualmente tienen un mismo significado pero están contenidas en los documentos escritas de formas diferentes. Se propone utilizar el algoritmo de n-gramas para solventar este problema.

Los métodos multivariados son utilizados para representar en mapas los resultados de la coocurrencia de palabras. Dentro de estos se pueden encontrar el Análisis de Componentes Principales (PCA), Escalamiento Multidimensional (MDS) y el agrupamiento.

El PCA es una técnica multivariable mediante la cual se analiza una tabla de datos que contienen observaciones cuantitativas realizadas a individuos.

Los objetivos del PCA son:

- Extraer la información más importante de la tabla de datos.
- Reducir el tamaño del conjunto de datos.
- Simplificar la descripción del conjunto de datos.
- Analizar la estructura y relación de las observaciones y de las variables.

2.1.2.3 Análisis de redes sociales (ARS)

El sistema debe permitir realizar este tipo de análisis reticular para reconocer los sistemas de relaciones presentes en una comunidad científica. Siendo capaz de encontrar las relaciones sociales existentes en los documentos científicos, representadas en los vínculos de coautoría, referenciales y de cooperación interinstitucional.

Estas relaciones en un plano general permiten reconocer en el ámbito científico tecnológico qué autores trabajan en qué áreas; cuáles son las comunidades estructuralmente fuertes en un tema determinado; oportunidades y amenazas que pueden afectar a una organización en áreas relacionadas con su campo de trabajo, actores centrales y periféricos dentro de las redes sociales que se construyen en una comunidad científica; posibles redes que permitan llevar a cabo proyectos conjuntos, rentabilidades, oportunidades de cambio científico tecnológico e innovaciones.

El análisis de redes parte de la base de que:

- Se puede pensar la sociedad en términos de estructuras.
- Dichas estructuras sociales se manifiestan en forma de relaciones entre actores sociales (sean éstos actores, grupos, organizaciones, clases o individuos).

Capítulo 2. Propuesta del Sistema

- Que los conjuntos de vínculos o de relaciones sociales forman redes.
- A su vez, según sea la posición que los diferentes actores intervinientes ocupan en dichas redes, van a definir sus valores, creencias y comportamientos.
- El principio de análisis no son los individuos ni los grupos, sino las relaciones y las redes de relaciones, los grupos surgen de las redes de relaciones y están cruzados por la pertenencia de sus miembros a distintas redes.

El análisis de redes sociales, toma elementos del álgebra de matrices y de la teoría de grafos para construir desde un conjunto delimitado de actores vinculados entre sí, una representación de las relaciones existentes.

2.1.3 Difusión de la información

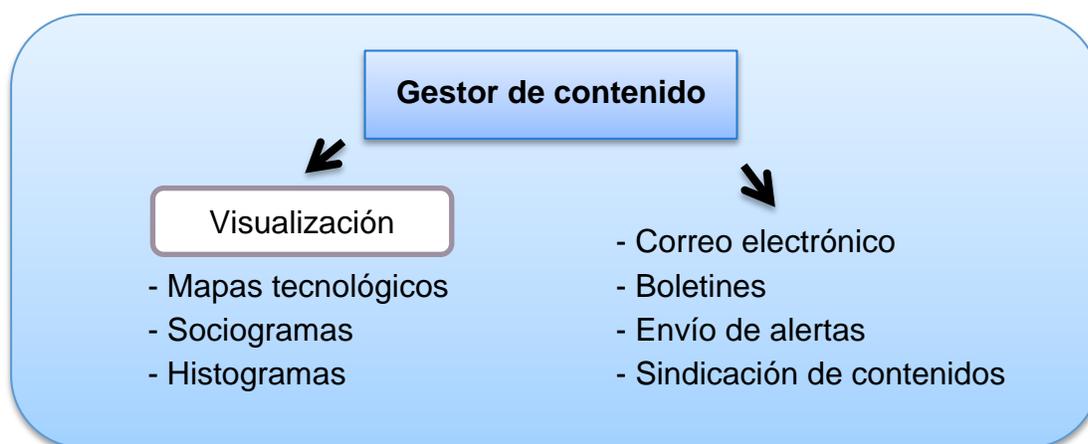


Figura 14. Difusión de la información del sistema.

Fuente: Elaboración propia

Como se puede observar en la figura anterior se utilizará un gestor de contenidos, el cual permitirá publicar boletines informativos, informes de vigilancia tecnológica, enviar por correo electrónico alertas a los usuarios personalizados por ellos mismos, la sindicación de contenidos para seguir algún tema de interés. De otro modo el sistema contendrá un módulo visualización el cual es el encargado de vincular los resultados de análisis de redes sociales con la extracción de palabras claves para presentar en un mapa las relaciones cognitivas y las relaciones sociales de tal manera que los investigadores puedan analizarlos y arribar a conclusiones.

Este módulo tendrá como resultado la modelación de mapas tecnológicos⁷, histogramas⁸ y sociogramas⁹, el objetivo es utilizar técnicas principalmente para visualizar y obtener

⁷ Son representaciones visuales del estado de la tecnología en un ámbito o área determinados, obtenidos a partir del tratamiento de la

Capítulo 2. Propuesta del Sistema

métricas de los resultados previamente obtenidos como pueden ser reducción de dimensionalidad, escalamiento multidimensional, agrupamiento, modelos gráficos probabilísticos y visualización de grafos.

2.2 Propuesta de Sistema Gestor de Contenidos (CMS)

Un CMS, brinda la posibilidad a un editor de crear, clasificar y publicar cualquier tipo de contenido en una página web.

James Robertson propone una división de la funcionalidad de los sistemas de gestión de contenidos en cuatro categorías: creación de contenido, gestión de contenido, publicación y presentación.

En el presente trabajo se utilizarán las categorías de publicación y presentación que brindan los CMS para la difusión de la información.

Publicación: *“una página aprobada se publica automáticamente cuando llega la fecha de publicación, y cuando caduca se archiva para futuras referencias. En su publicación se aplica el patrón definido para toda la web o para la sección concreta donde está situada, de forma que el resultado final es un sitio web con un aspecto consistente en todas sus páginas. Esta separación entre contenido y forma permite que se pueda modificar el aspecto visual de un sitio web sin afectar los documentos ya creados y libera a los autores de preocuparse por el diseño final de sus páginas.” [28]*

Presentación: *“el sistema se encarga de gestionar muchos otros aspectos como son los menús de navegación o la jerarquía de la página actual dentro de la web, añadiendo enlaces de forma automática. También gestiona todos los módulos, internos o externos, que incorpore al sistema. Así por ejemplo, con un módulo de noticias se presentarían las novedades aparecidas en otro web, con un módulo de publicidad se mostraría un anuncio o*

información contenida en bases de datos de patentes y artículos, (Escorsa, Rodríguez y Maspons, 2000). Los mapas presentan gráficamente, de forma sintética, las tecnologías en que se ha investigado más y, en consecuencia, publicado y patentado más en un período determinado. Permiten también detectar aquellas tecnologías emergentes que están experimentando una rápida expansión mediante la comparación con mapas correspondientes a períodos anteriores.

<http://www.revistaespacios.com/a00v21n02/41002102.html>

⁸ Es un resumen gráfico de los valores producidos por las variaciones de una determinada característica, representando la frecuencia con que se presentan distintas categorías dentro de dicho conjunto.

<http://www.fundibeg.org/opencms/export/sites/default/PWF/downloads/gallery/methodology/tools/histograma.pdf>

⁹ “La técnica del sociograma consiste en representar gráficamente las relaciones interpersonales en un grupo de individuos mediante un conjunto de puntos (los individuos) conectados por una o varias líneas (las relaciones interindividuales). [...] (En el sociograma) las líneas empiezan a ser relaciones sociales de cualquier tipo y los puntos entidades sociales, que no se identifican necesariamente con los individuos” <http://webcasus.usal.es/edenred/documentos/Mapas Sociales Pedro Mart n .rtf>

Capítulo 2. Propuesta del Sistema

mensaje animado, y con un módulo de foro se podría mostrar, en la página principal, el título de los últimos mensajes recibidos. Todo eso con los enlaces correspondientes y, evidentemente, siguiendo el patrón que los diseñadores hayan creado". [28]

Estas funcionalidades brindan la respuesta al por qué utilizar un sistema de gestión de contenido, además que el uso de un CMS tiene implícito una serie de **ventajas** en el campo de la publicación de información como:

- Inclusión de nuevas funcionalidades en la web
- Mantenimiento de gran cantidad de páginas
- Reutilización de objetos o componentes
- Páginas interactivas
- Cambios en el aspecto de la web
- Consistencia de la web
- Control de acceso

Según el tipo de licencia los CMS presentan dos clasificaciones: comerciales y de código abierto.

Los primeros son armas de las grandes compañías de diseño y desarrollo web, su uso y soporte técnico tiene un elevado costo. Entre estos se encuentran: CoreMedia CMS, PipePS, NUKE ET.

Los de código abierto son desarrollados por grandes grupos de personas, empresas y comunidades de desarrollo; las posibilidades de personalización del producto son elevadas, el código fuente puede ser modificado permitiendo la corrección de errores, y el desarrollo de nuevas funcionalidades.

Existen varios CMS basados en código abierto como: ASP Nuke, Drupal, Plone. Para el sistema se propone el CMS Drupal teniendo en cuenta los siguientes elementos:

- Brinda a los programadores expertos una plataforma altamente flexible para soportar sus aplicaciones a través del desarrollo de *plugins* o módulos que se integran con el sistema, de esta forma el desarrollador puede hacer uso provechoso de las funcionalidades que brinda como plataforma.
- Conveniente para este trabajo que se desea implementar, dado que posee varios módulos que serían de gran ayuda. Por ejemplo, el módulo Workflow (Flujo de

Capítulo 2. Propuesta del Sistema

trabajo) es implementado para al traspaso de información antes de ser publicada a través de un flujo de trabajo manejado por diferentes roles asignados a varias personas. Con este módulo se asegura que la información sólo pueda ser vista por un rol específico en un momento determinado.

- Drupal es de código abierto con una comunidad a su alrededor muy extensa, de variados idiomas y conocimientos, esto lo ha hecho crecer y posicionarse encima de los demás CMS, Drupal es libre para descargarlo de Internet y usarlo.
- La herramienta, va orientada al grupo GIC el cual usa el CMS Drupal en el desarrollo de sus aplicaciones informáticas, ejemplo el observatorio tecnológico en construcción, por tal motivo para la herramienta se sigue la misma política.

2.3 Arquitectura

El sistema será implementado usando el CMS Drupal, por lo que la arquitectura se hereda del mismo.

La arquitectura de CMS Drupal utiliza el patrón Modelo Vista Controlador (MVC). Este patrón de arquitectura de software separa los datos de una aplicación, la interfaz de usuario, y la lógica de control en tres componentes distintos. El patrón MVC se ve frecuentemente en aplicaciones web, donde la vista es la página HTML y el código que provee de datos dinámicos a la página; el modelo es el sistema de gestión de base de datos y la Lógica de negocio; y el controlador es el responsable de recibir los eventos de entrada desde la vista.

Capítulo 2. Propuesta del Sistema

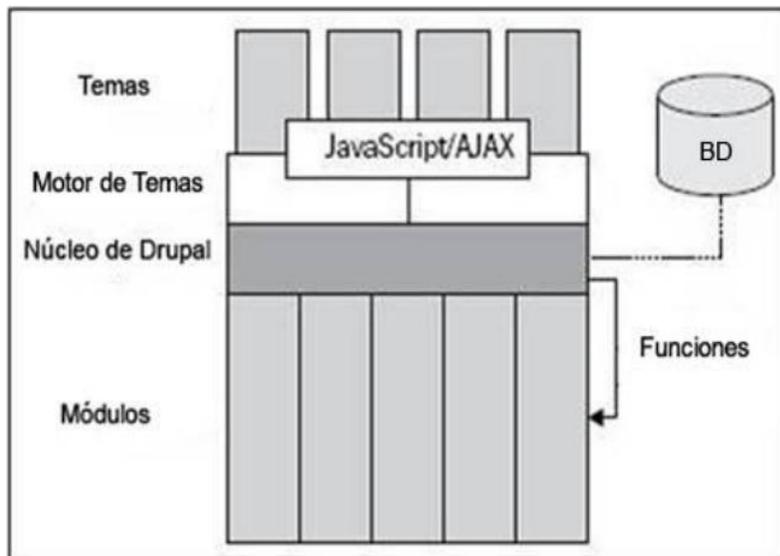


Figura 15. Arquitectura de software del CMS Drupal.

En la figura anterior se muestra la estructura en capas del CMS Drupal, donde la vista son los temas, el motor de temas y javascript/ajax; el modelo es la base de datos (DB), y el controlador es el núcleo de Drupal y los módulos; donde funciones, son los métodos para facilitar la comunicación entre el núcleo de Drupal y los módulos.

2.4 Propuesta de tecnologías para la implementación del sistema

2.4.1 Lenguajes de programación

A partir de las características que presenta un sistema web, se ha decidido adoptar una arquitectura cliente - servidor.

En el modelo CLIENTE/SERVIDOR podemos encontrar las siguientes características [29]:

- El cliente y el servidor pueden actuar como una sola entidad y también pueden actuar como entidades separadas, realizando actividades o tareas independientes.
- Las funciones de cliente y servidor pueden estar en plataformas separadas, o en la misma plataforma.
- Un servidor da servicio a múltiples clientes en forma concurrente.
- Cada plataforma puede ser escalable independientemente. Los cambios realizados en las plataformas de los clientes o de los servidores, ya sean por actualización o por reemplazo tecnológico, se realizan de una manera transparente para el usuario final.

Capítulo 2. Propuesta del Sistema

- La interrelación entre el hardware y el software están basados en una infraestructura poderosa, de tal forma que el acceso a los recursos de la red no muestra la complejidad de los diferentes tipos de formatos de datos y de los protocolos.

Actualmente existen un gran número de lenguajes de programación que permiten desarrollar una aplicación con las características planteadas anteriormente, es decir, lenguajes basados en la tecnología web; los cuales pueden dividirse en dos grandes grupos: lenguajes del lado del cliente y lenguajes del lado del servidor.

2.4.1.1 Lenguajes del lado del cliente. [30]

HTML es el lenguaje utilizado para la creación de páginas Web, significa “HyperText Mark-Up Language”, en español, “Lenguaje para el Formato de Documentos de Hipertexto”. Es decir, los documentos HTML no son de texto normal, sino de hipertexto ya que en el propio documento aparecen enlaces a otros.

El lenguaje HTML es un estándar reconocido en todo el mundo y cuyas normas define un organismo sin ánimo de lucro llamado World Wide Web Consortium, más conocido como W3C. Como se trata de un estándar reconocido por todas las empresas relacionadas con el mundo de Internet, una misma página HTML se visualiza de forma muy similar en cualquier navegador de cualquier sistema operativo.

El propio W3C define el lenguaje HTML como "un lenguaje reconocido universalmente y que permite publicar información de forma global". Desde su creación, el lenguaje HTML ha pasado de ser un lenguaje utilizado exclusivamente para crear documentos electrónicos a ser un lenguaje que se utiliza en muchas aplicaciones electrónicas como buscadores, tiendas online y banca electrónica.

Todo documento HTML requiere expresar correctamente una serie de etiquetas, y todo elemento de este tipo de documento vendrá dado por la siguiente estructura: <etiquetaInicial>Texto</etiquetaInicial> como se puede ver en la siguiente figura.

Capítulo 2. Propuesta del Sistema

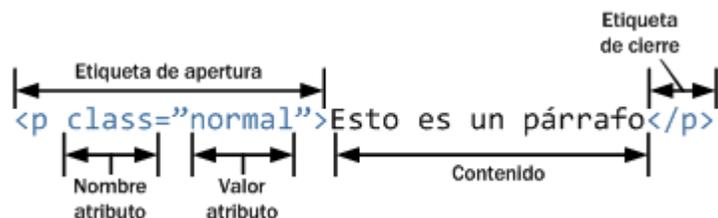


Figura 16. Esquema de las partes que componen un elemento HTML.

Fuente: LibrosWeb.Elementos HTML. http://www.librosweb.es/xhtml/capitulo2/elementos_html.html.

JavaScript es un lenguaje de programación que se utiliza principalmente para crear páginas web dinámicas.” [31]

“Una página web dinámica es aquella que incorpora efectos como texto que aparece y desaparece, animaciones, acciones que se activan al pulsar botones y ventanas con mensajes de aviso al usuario. Es un lenguaje de programación interpretado, por lo que no es necesario compilar los programas para ejecutarlos. Los programas escritos con JavaScript se pueden probar directamente en cualquier navegador sin necesidad de procesos intermedios.” [31]

“A pesar de su nombre, JavaScript no guarda ninguna relación directa con el lenguaje de programación Java. Legalmente, JavaScript es una marca registrada de la empresa Sun Microsystems.” [31]

CSS es un lenguaje de hojas de estilos creado para controlar el aspecto o presentación de los documentos electrónicos definidos con HTML y XHTML. CSS es la mejor forma de separar los contenidos y su presentación y es imprescindible para crear páginas web complejas.” [32]

“Separar la definición de los contenidos y la definición de su aspecto presenta numerosas ventajas, ya que obliga a crear documentos HTML/XHTML bien definidos y con significado completo (también llamados “documentos semánticos”). Además, mejora la accesibilidad del documento, reduce la complejidad de su mantenimiento y permite visualizar el mismo documento en infinidad de dispositivos diferentes”. [32]

2.4.1.2 Lenguajes del lado del servidor

PHP (acrónimo de PHP: Hypertext Preprocessor), es un lenguaje interpretado de alto nivel embebido en páginas HTML y ejecutado en el servidor”. [33]

Capítulo 2. Propuesta del Sistema

“Una de sus características más potentes es su soporte para gran cantidad de bases de datos. Entre estas pueden mencionarse InterBase, mSQL, MySQL, Oracle, Informix, PostgreSQL, entre otras.” [33]

“Como producto de código abierto, goza de la ayuda de un gran grupo de programadores, permitiendo que los fallos de funcionamiento se encuentren. El código se pone al día continuamente con mejoras y extensiones de lenguaje para ampliar sus capacidades.” [33]

2.4.2 Gestor de base de datos

Drupal, fue el CMS propuesto para el desarrollo de la herramienta. Los desarrolladores de este CMS concibieron su modelo de acceso a datos para utilizar varios sistemas gestores de base de datos dentro de estos MySQL y PostgreSQL. Teniendo en cuenta esto a continuación se realiza una valoración de las características fundamentales de estos gestores para seleccionar el más conveniente a utilizar en el sistema.

“MySQL es un sistema de administración de bases de datos relational (RDBMS). Se trata de un programa capaz de almacenar una enorme cantidad de datos de gran variedad y de distribuirlos para cubrir las necesidades de cualquier tipo de organización, desde pequeños establecimientos comerciales a grandes empresas y organismos administrativos.” [34]

Características:

- El principal objetivo de MySQL es la velocidad.
- Soporta gran cantidad de tipos de datos para las columnas.
- Gran portabilidad entre sistemas, puede trabajar en distintas plataformas y sistemas operativos.
- Cada base de datos cuenta con 3 archivos: uno de estructura, uno de datos y uno de índice y soporta hasta 32 índices por tabla.
- Aprovecha la potencia de sistemas multiproceso, gracias a su implementación multihilo.
- Flexible sistema de contraseñas y gestión de usuarios, con un muy buen nivel de seguridad en los datos.
- El servidor soporta mensajes de error en distintos lenguajes.

Capítulo 2. Propuesta del Sistema

Ventajas: [35]

- Velocidad al realizar las operaciones, lo que le hace uno de los gestores con mejor rendimiento.
- Bajo costo en requerimientos para la elaboración de bases de datos, ya que debido a su bajo consumo puede ser ejecutado en una máquina con escasos recursos sin ningún problema.
- Facilidad de configuración e instalación.
- Soporta gran variedad de Sistemas Operativos.
- Baja probabilidad de corromper datos, incluso si los errores no se producen en el propio gestor, sino en el sistema en el que está.
- Conectividad y seguridad.

Desventajas: [35]

- Un gran porcentaje de las utilidades de MySQL no están documentadas.
- No es intuitivo, como otros programas.

“PostgreSQL es un sistema de gestión de bases de datos relacionales Open Source (de código abierto), gratuito y que al tener licencia de tipo BSD, nos permite manejar libremente el código fuente del gestor de bases de datos PostgreSQL, mejorando u optimizando su código. Incluso se permite redistribuirlo como producto comercial y combinarlo con herramientas de licencia propietaria.” [36]

Ventajas: [36]

- Posee manejo y control de transacciones para asegurar la consistencia de los datos.
- Soporta los tipos de datos, cláusulas, funciones y comandos de tipo estándarSQL92/SQL99 y extendidos propios de PostgreSQL. Los tipos de datos internos han sido mejorados incluyendo nuevos tipos.
- PostgreSQL puede operar sobre distintas plataformas incluyendo Linux, Unix, MacOSX, Solaris y Windows.
- Posee un buen sistema de seguridad mediante la gestión de usuarios, grupos de usuarios, permisos y contraseñas.

Capítulo 2. Propuesta del Sistema

- Posee una gran capacidad de almacenamiento.
- Posee ciertas características de orientación a objetos, como la herencia entre tablas.
- Tiene una buena escalabilidad ya que es capaz de ajustarse al número de CPUs y a la cantidad de memoria que posee el sistema de forma óptima, soportando una mayor cantidad de peticiones simultáneas a la base de datos de manera correcta.

Desventajas: [37]

- En comparación con MySQL es más lento en inserciones y actualizaciones, ya que cuenta con cabeceras de intersección que no tiene MySQL.
- Soporte en línea: Hay foros oficiales, pero no hay una ayuda obligatoria.
- La sintaxis de algunos de sus comandos o sentencias no es intuitiva.

Valoración de los gestores de bases de dato

Como gestor se selecciona PostgreSQL debido a que posee una gran escalabilidad; es capaz de ajustarse al número de CPU y a la cantidad de memoria que posee el sistema, haciéndole capaz de soportar una mayor cantidad de peticiones simultáneas de manera correcta que MySQL. A pesar de ser más lento que MySQL en las consultas, cuando existe una enorme cantidad de datos puede llegar a ser igual o más rápido que MySQL. Es un sistema que asegura la consistencia de la base de datos.

2.4.3 Servidores web

“Según el ranking de servidores más usados desde 1995 hasta octubre de 2011 consultado en netcraft.com¹⁰, Apache sigue siendo el número uno en mundo seguido por IIS de Microsoft.” [38]

Se decide utilizar apache ya que:

- Presenta licencia de software libre y de código abierto sin embargo Internet Information Services (IIS) su licencia es de software propietario.
- Apache es multiplataforma a diferencia de IIS que funciona solo sobre la plataforma Windows, lo que es una limitante.

¹⁰ Netcraft es una compañía de servicios de Internet con sede en Bath, Inglaterra. Proporcionar datos de la investigación y análisis sobre muchos aspectos de Internet. <http://news.netcraft.com/about-netcraft/>

Capítulo 2. Propuesta del Sistema

- Apache tiene una alta comunidad de usuarios lo que posibilita que se le dé solución tanto a errores del propio servidor como a errores de configuración.
- Apache da importancia a la estabilidad y seguridad de su servidor. El rendimiento es una segunda prioridad sin embargo para ISS el rendimiento ha sido su prioridad.

Consideración de las tecnologías propuestas para el desarrollo

Como resultado del proceso investigativo de las principales herramientas y tecnologías, se realiza la propuesta de las mismas, teniendo como principal criterio que estas pertenezcan a la familia del software libre. A continuación se presenta la selección realizada:

Se propone HTML, CSS versión 2, JavaScript como lenguajes del lado del cliente y PHP versión 5.3.10 como lenguaje de programación por la amplia documentación que existe en la universidad y en Internet, la flexibilidad y la claridad de su código abierto, además se ajusta a los requisitos de la aplicación. Con PHP como punto de partida es necesario integrar a otras herramientas incluyendo así a Apache como servidor web y PostgreSQL como gestor de bases de datos conformando una robusta plataforma de trabajo para el desarrollo de aplicaciones web.

2.5 Roles para el sistema de vigilancia

En la tesis de pre-grado de Linnet Ivet Bello Leyva, precedente investigación, como ya se ha planteado anteriormente se definen los roles para llevar a cabo el proceso de VT en una institución. Por tal motivo en el presente trabajo se asumen los mismos, siendo mencionados a continuación:

“Buscadores/Captadores de información: dentro de la estructura organizativa es preciso contar con recursos humanos capacitados para la búsqueda eficiente de información estratégica, con experiencia en el dominio de las herramientas existentes, bases de datos, etcétera. Este rol aporta un alto valor añadido en cuanto a la selección de fuentes y primer filtrado de información.” [15].

“Red de expertos internos y externos: este colectivo aporta un análisis crítico y consistente de la base de información recopilada. Conjugando su experiencia y conocimientos extraen el valor añadido estratégico para la organización. En el caso de expertos externos son determinantes los niveles de confianza y confidencialidad.” [15]

“Decisores: este rol directivo soporta el proceso de toma de decisiones, protagonizando, por tanto, el papel de clientes del sistema de gestión de la información. Desde este

Capítulo 2. Propuesta del Sistema

colectivo se establecen las necesidades a modo de factores clave de vigilancia recibiendo las respuestas e información oportunas para la acción estratégica.” [15]

“**Responsable/animador de la gestión de la información:** la figura del animador resulta fundamental para otorgar dinamismo al sistema de gestión de la información. En este sentido, este rol identifica a los participantes clave en el proceso y realiza funciones de motivación, coordinación de las diferentes áreas y miembros de la organización. Así, el animador transmite la apetencia por el reto estratégico que supone la consecución de una estructura organizativa inteligente.” [15]

2.6 Modelo de dominio

Durante el análisis de la problemática planteada no fue posible identificar procesos del negocio visiblemente ni los roles que participaban en el mismo. Siendo necesario representar la situación mediante un modelo de dominio.

Un modelo de dominio es un artefacto de la disciplina de análisis, construido con las reglas de UML durante la fase de concepción, presentado como uno o más diagramas de clases que contiene conceptos del sistema de software.

El proceso para su elaboración tiene tres pasos:

- Identificar las clases conceptuales
- Dibujarlas en un diagrama de clases
- Añadir relaciones y atributos

Capítulo 2. Propuesta del Sistema

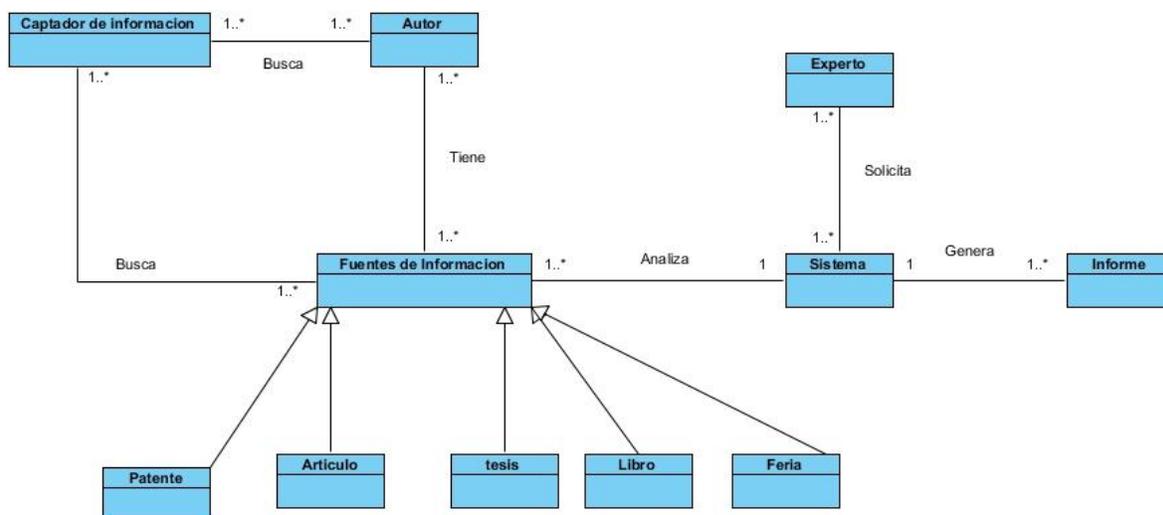


Figura 17. Modelo de dominio del sistema de vigilancia tecnológica.

El modelo de dominio correspondiente al sistema propuesto (Figura 17) cuenta con las siguientes clases:

Captador de información: Actor que gestiona la información de las fuentes.

Fuentes de información: Contiene todas las fuentes de información.

Patente: Contiene los metadatos de las patentes.

Artículo: Contiene los metadatos de los artículos.

Tesis: Contiene los metadatos de las tesis.

Libro: Contiene los metadatos de los libros.

Ferías: Contiene los metadatos de las ferias.

Experto: Rol encargado de interactuar con el sistema.

Sistema: Módulos encargados del procesamiento, análisis y visualización de la información.

Informe: Contiene informes de vigilancia tecnológica.

Autor: Contiene los metadatos de los autores.

2.7 Especificación de Requisitos

“Los requisitos son capacidades o condiciones que debe tener presente un sistema para satisfacer las necesidades de un usuario o resolver un problema. Constituyen la base del proceso de desarrollo del software en cuestión. ”

Capítulo 2. Propuesta del Sistema

Los requerimientos deben ser: [39]

- Especificado por escrito: Como todo contrato o acuerdo entre dos partes.
- Posible de probar o verificar. Si un requerimiento no se puede comprobar, entonces ¿cómo se sabe si se cumplió con él o no?
- Conciso: Un requerimiento es conciso si es fácil de leer y entender. Su redacción debe ser simple y clara para aquellos que vayan a consultarlo en un futuro.
- Completo: Un requerimiento está completo si no necesita ampliar detalles en su redacción, es decir, si se proporciona la información suficiente para su comprensión.
- Consistente: Un requerimiento es consistente si no es contradictorio con otro requerimiento.
- No ambiguo: Un requerimiento no es ambiguo cuando tiene una sola interpretación. El lenguaje usado en su definición, no debe causar confusiones al lector.

Para identificar los requerimientos del presente sistema se analizaron otros similares ya desarrollados. Esta técnica es conocida como Sistemas Existentes y permite extraer información al interactuar con las interfaces de usuario, observando el tipo de información que se maneja y cómo es manejada, así como las distintas salidas que los sistemas producen.” [40]

2.7.1 Requisitos Funcionales

“Los requerimientos funcionales son los que definen las funciones que el sistema será capaz de realizar, describen las transformaciones que el sistema realiza sobre las entradas para producir salidas. Es importante que se describa el ¿Qué? y no el ¿Cómo? se deben hacer esas transformaciones. Estos requerimientos al tiempo que avanza el proyecto de software se convierten en los algoritmos, la lógica y gran parte del código del sistema.” [39]

A continuación se describen los requisitos funcionales que debe cumplir el presente sistema:

R-1: Gestionar fuentes.

R-1.1: Insertar metadatos de fuentes: El sistema debe permitir insertar los metadatos de las fuentes de información según el tipo.

R-1.2: Modificar fuentes: Debe permitir cambiar lo metadatos de las fuentes según sea necesario.

Capítulo 2. Propuesta del Sistema

R-1.3: Eliminar Fuentes: Eliminar una o varias fuentes

R-2: Gestionar autor

R-2.1: Insertar autor: El sistema debe permitir insertar los metadatos de un autor.

R-2.2: Modificar autor: Debe permitir cambiar lo metadatos de un autor.

R-2.3: Eliminar autor: Eliminar un autor.

R-3: Normalizar palabras claves: El sistema normalizará las palabras claves de las patentes y artículos.

R-4: Contar palabra: El sistema debe determinar la frecuencia de palabras clave que se repiten en los documentos, que se encuentran en los campos título, abstracts, descriptores y de los identificadores.

R-5: Contar palabras claves por año: El sistema mostrará gráficamente la frecuencia de palabras claves por año.

R-6: Contar autores por palabras clave y año: El sistema mostrará gráficamente la frecuencia de autores por palabras claves de un año determinado.

R-7: Buscar coocurrencia de palabras: El sistema debe ser capaz de encontrar coocurrencia de palabras en los campos de los documentos.

R-8: Modelar relaciones sociales: Esto permite modelar los documentos según las relaciones sociales como son autores o instituciones que intervienen en el mismo.

2.7.2 Requisitos no funcionales

“Los requerimientos no funcionales tienen que ver con características que de una u otra forma puedan limitar al sistema, como por ejemplo, el rendimiento (en tiempo y espacio), interfaces de usuario, fiabilidad (robustez del sistema, disponibilidad de equipo), mantenimiento, seguridad, portabilidad, estándares, etc.” [39]

El sistema en cuestión debe respetar los siguientes requisitos no funcionales:

Usabilidad

- Utilizar nombres sugerentes para lograr que el usuario encuentre lo que busca en el menor tiempo posible.
- Mostrar textos explicativos al pasar el puntero por encima de los íconos.
- Utilizar estándares para descripción de las fuentes de información.

Capítulo 2. Propuesta del Sistema

Portabilidad:

- Al sistema se debe acceder desde cualquier sistema operativo.
- Al sistema se debe acceder desde cualquier navegador.

Seguridad y privacidad:

- Confidencialidad:

La información manejada por el sistema debe estar protegida ante el acceso no autorizado y la divulgación. El acceso al sistema se restringirá a los usuarios de acuerdo al rol que desempeñen. El sistema debe garantizar que datos sensibles no viajen en texto plano por la red.

- Integridad:

La información manejada por el sistema será objeto de cuidadosa protección contra la corrupción y estados inconsistentes, de la misma forma será considerada igual la fuente o autoridad de los datos.

- Eficiencia:

La herramienta propuesta debe ser rápida y el tiempo de respuesta debe ser el mínimo posible (6 segundos máximo), adecuado a la rapidez con que el usuario requiere la respuesta a su acción.

Restricciones de diseño e implementación:

- El sistema se implementará con tecnología PHP 5.3.10.

Interfaces de hardware:

PC Servidor

- Microprocesador 2.66 GHz
- 4 GB de RAM.
- Espacio disponible en disco: 1 TB.

PC Cliente

- Microprocesador 2.66 GHz.
- 512 MB de RAM.

Requisitos de Licencia:

Capítulo 2. Propuesta del Sistema

- Todas las tecnologías a utilizar deben poseer licencias libres.

2.8 Modelo de Casos de Uso del Sistema

El modelo de casos de uso del sistema representa las funcionalidades y acciones de los usuarios sobre el sistema a desarrollar a través de los casos de uso y actores que pueden ser tanto personas como máquinas cuyo rol o autoridad le permiten iniciarlo.

2.8.1 Patrones de Casos de Uso

Patrones de Casos de Uso: Son comportamientos que deben existir en el sistema, ayudan a describir qué es lo que el sistema debe hacer, es decir, describen el uso del sistema y cómo este interactúa con los usuarios. Estos patrones son utilizados generalmente como plantillas que describen como debería ser estructurados y organizados los casos de uso. Son patrones que capturan mejores prácticas para modelar casos de uso. [41]

Existen diferentes patrones aplicados a casos de uso entre los que podemos destacar: [41]

Concordancia (Commonality): Este patrón toma una subsecuencia de acciones que estén en diferentes partes del flujo de casos de uso y la define por separado.

Concordancia: Re-uso (Commonality: Reuse): El patrón Re-uso es un patrón de estructura que consiste en tres casos de uso. El primero llamado “Sub-secuencia Común”, modela la secuencia de acciones que aparecen en múltiples casos de uso del modelo. Los otros dos casos de uso comparten de esta sub-secuencia común de acciones (dos es la menor cantidad que puede existir).

Concordancia: Adición (Commonality: Addition): En el caso de este patrón alternativo, la subsecuencia común de casos de uso, extiende los casos de uso compartiendo la subsecuencia de acciones. Los otros casos de uso modelan el flujo que será expandido con la subsecuencia. Este patrón es preferible usarlo cuando otros casos de uso se encuentran propiamente completos, o sea, que no requieren de una subsecuencia común de acciones para modelar los usos completos del sistema.

Extensión Concreta (Concrete Extension or Inclusion: Extension): Extensión Concreta es un patrón de estructura. Patrón que consiste en dos casos de uso y una relación de extensión. El caso de uso extendido es concreto, es decir, puede ser instanciado por su cuenta como por el caso de uso base. Este patrón es aplicable cuando un flujo puede

Capítulo 2. Propuesta del Sistema

extender el flujo de otro caso de uso, lo que significa que puede ocurrir el proceso del caso de uso base, o puede ocurrir el del caso de uso base con su caso de uso extendido.

Inclusión Concreta (Concrete Extension or Inclusion: Inclusion): Inclusión Concreta es un patrón de estructura. Consiste en dos casos de uso y una relación de inclusión entre el caso de uso base y el caso de uso incluido. Este último puede ser instanciado por sí solo. El caso de uso base puede ser concreto o abstracto. Se utiliza este patrón cuando un flujo de datos puede ser incluido en el flujo de otro caso de uso y también puede ejecutarse por sí solo.

CRUD (Creating, Reading, Updating, Deleting): Es un patrón de estructura que se basa en la fusión de casos de uso simples para formar una unidad conceptual.

Múltiples actores (Multiple Actors): Captura la concordancia entre actores manteniendo roles separados.

Roles diferentes (Distinct Roles): Consiste en un caso de uso y por lo menos dos actores. Este patrón se usa cuando los actores interactúan de forma diferentemente sobre el caso de uso.

Roles comunes (Commons Roles): Puede suceder que los dos actores jueguen el mismo rol sobre el caso de uso. Este rol es representado por otro actor, heredado por los actores que comparten este rol. Es aplicable cuando, desde el punto de vista del caso de uso, solo exista una entidad externa interactuando con cada una de las instancias del caso de uso.

En la modelación del caso de uso del sistema se utilizó el patrón **CRUD**.

Actores del sistema

Los actores del presente sistema son los siguientes:

Tabla 1. Actores

Actores	Descripción
Captador	Rol encargado de gestionar las fuentes de información y los autores
Experto	Rol encargado de operar con las funcionalidades de análisis, procesamiento y visualización del sistema

Capítulo 2. Propuesta del Sistema

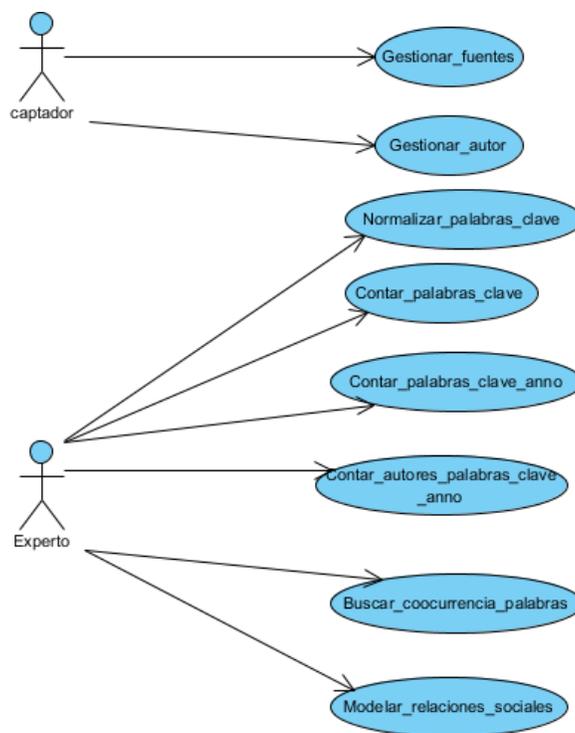


Figura 18. Diagrama de caso de uso del sistema.

2.8.2 Descripción de casos de uso del sistema

Los casos de uso identificados describen la secuencia de interacciones que se producen entre un actor y el sistema. Estableciendo un acuerdo entre clientes y desarrolladores sobre las condiciones y cualidades que deben cumplir las funcionalidades.

A continuación se presenta la descripción del caso de uso Buscar coocurrencia de palabras. Para el estudio de las demás descripción remitirse al Anexo 3: Descripciones de los casos de usos.

Tabla 2. Descripción del caso de uso Buscar coocurrencia de palabras.

Caso de Uso:	Buscar coocurrencia de palabras
Actores:	experto
Resumen:	El caso de uso comienza cuando el actor selecciona la funcionalidad Buscar coocurrencia de palabras.
Precondiciones:	Los campos deben estar llenos de las fuentes de patentes y artículos científicos de revista. Deben estar normalizadas las palabras claves.
Referencias:	RF-7
Prioridad:	Alta
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
1 Selecciona la opción	1.1 Se extraen la raíz de las palabras claves del corpus.

Capítulo 2. Propuesta del Sistema

buscar coocurrencia de palabras	
	2 Crea matrices binarias de ocurrencia o no de las palabras en los documentos, matriz (documentos, raíz de palabras claves).
	2.1 Busca en cada documento la existencia de las palabras claves usando la raíz y va llenado la matriz.
	3 Elimina las palabras claves que tenga frecuencia menor a un número predeterminado.
	3.1 Crea matriz de número ocurrencia y coocurrencia de las palabras claves en los documentos, sería matriz (palabras claves, palabras claves) partiendo de la matriz anterior.
	4 Elimina las coocurrencias menores a un número predeterminado.
	4.1 Determina el número de palabras máximo y mínimo por clúster.
	5 Calcula índice de Jaccard para hallar la similitud de los documentos (Ver anexo 4).
	5.1 Modela mapas tecnológicos a través de técnicas de visualización como PCA, MDS y agrupamiento.
6 Selecciona la opción de guardar	6.1 Muestra pantalla(año, descripción).
7 Inserta los datos y acepta	7.1 Guarda los datos.
Flujos Alternos	
7. a Selecciona la opción Cancelar	
7.a.1 Cancela la opción de guardar	7.a.2 Muestra pantalla de mapa tecnológico.
Poscondiciones	Queda construido las relaciones cognitivas a través de mapas tecnológicos con diferentes técnicas de visualización.

2.9 Análisis y diseño

El análisis y diseño del sistema permitirá entender el negocio, definir los requisitos funcionales y no funcionales, modelar los diagramas clases de análisis, diseño, de tal manera que para el desarrollador resulte más asequible su implementación.

2.9.1 Modelo de Análisis

Los diagramas de clases del análisis representan las relaciones entre los actores y el sistema, que se muestra seccionado en clases: interfaz, controladoras y entidades; que pueden ser la abstracción de una o varias clases o subsistemas del diseño del sistema.

Tabla 3. Descripción de las clases del análisis

Estereotipo	Descripción
 <p style="text-align: center;">CI</p>	Clases Interfaz: Se utilizan para modelar la interrelación entre el sistema y sus actores (es decir usuarios y sistemas externos). Las clases interfaz modelan las partes del

Capítulo 2. Propuesta del Sistema

<p>Estereotipo de las clases interfaz.</p>	<p>sistema que dependen de sus actores lo que implica que clarifiquen y reúnan los requisitos en los límites del sistema.</p>
<p style="text-align: center;">  CC </p> <p>Estereotipo de las clases control.</p>	<p>Clases Control: Representan la coordinación, secuencia, transacciones y control de objetos, se usan con frecuencia para encapsular el control de un caso de uso en concreto.</p>
<p style="text-align: center;">  CE </p> <p>Estereotipo de las clases entidad.</p>	<p>Clases Entidad: Se utilizan para modelar información que posee una vida larga y que es a menudo persistente. En la mayoría de los casos se derivan directamente de una clase entidad del negocio. [42]</p>

A continuación se muestra el diagrama de clase del análisis correspondiente al caso de uso Buscar coocurrencia de palabras, para el estudio de los demás diagramas remitirse al Anexo 5: Diagramas de clases del análisis.

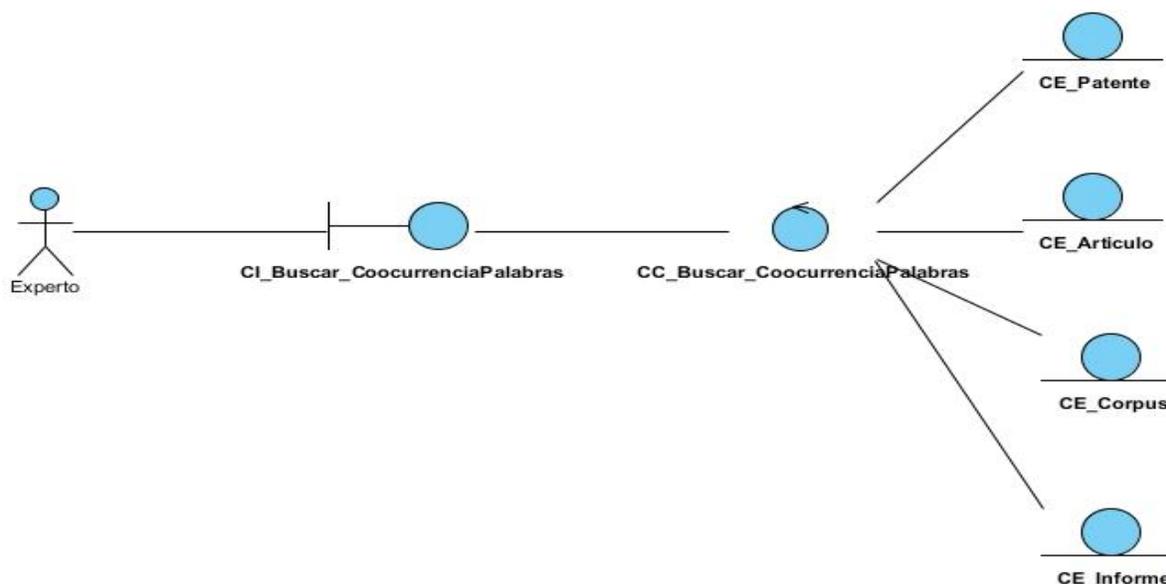


Figura 19. Diagrama de clase del análisis Buscar coocurrencia de palabras.

Capítulo 2. Propuesta del Sistema

2.9.1.1 Diagramas de interacción del análisis

Los diagramas de interacción representan la interacción entre los actores y el sistema para obtener determinado resultado. La metodología RUP propone su realización a través de los diagramas de colaboración o los diagramas de secuencia.

Los diagramas de secuencia identifican la secuencia y orden cronológico de acciones. Estos diagramas contribuyen a la descripción de la dinámica del sistema en términos de la interacción entre sus objetos.

Por su parte los diagramas de colaboración permiten identificar los requisitos funcionales y las responsabilidades sobre los objetos. Muestran las relaciones entre objetos acompañadas de mensajes que indican el objetivo de la misma.

A continuación se muestran el diagrama de colaboración del caso de uso Buscar coocurrencia de palabras, para consultar los demás diagramas ver Anexo 6: Diagramas de colaboración.

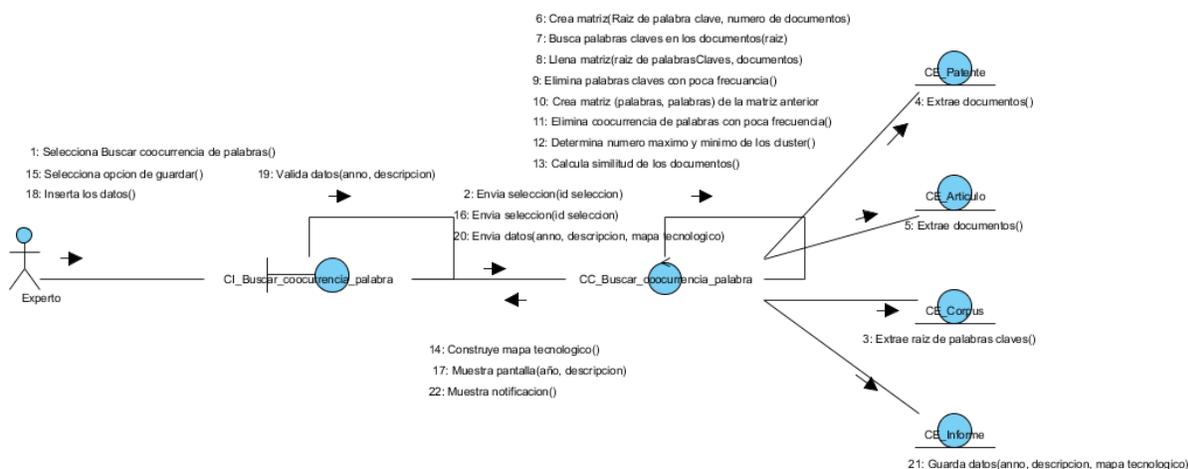


Figura 20. Diagrama de colaboración Buscar coocurrencia de palabras.

2.9.2 Modelo de diseño

El modelo de diseño es un modelo de objetos que describe la realización física de los casos de uso centrándose en cómo los requisitos funcionales y no funcionales, junto con otras restricciones relacionadas con el entorno de implementación, tienen impacto en las funcionalidades que se desean implementar.

Capítulo 2. Propuesta del Sistema

2.9.2.1 Patrones de diseño

Los patrones de diseño son la base para la búsqueda de soluciones a un problema de diseño o a problemas comunes en el desarrollo de software y otros ámbitos referentes al diseño de interacción o interfaces.

Los patrones de diseño tienen como ventajas que:

- Proponen una forma de reutilizar la experiencia de los desarrolladores.
- Están basados en la recopilación del conocimiento de los expertos en desarrollo de software.
- Es una experiencia real, probada y funciona. Es historia y ayuda a no cometer los mismos errores.

Entre los que más se utilizan se encuentran los patrones GRASP que son los patrones para la asignación de responsabilidades, donde se destacan por su uso en el diseño:

Experto: Asignar una responsabilidad al experto en información, la clase que cuenta con la información necesaria para cumplir la responsabilidad.

Alta cohesión: La cohesión es una medida de cuán relacionadas y enfocadas están las responsabilidades de una clase. Una alta cohesión caracteriza a las clases con responsabilidades estrechamente relacionadas que no realicen un trabajo enorme. Este patrón expresa que la información que almacena una clase debe de ser coherente y está en la mayor medida de lo posible relacionada con la clase.

Bajo acoplamiento: Es la idea de tener las clases lo menos ligadas entre sí posible. De tal forma que en caso de producirse una modificación en alguna de ellas, tenga la mínima repercusión posible en el resto de las clases, potenciando la reutilización, y disminuyendo la dependencia entre clases.

Creador: Este patrón como su nombre lo indica es el que crea, el guía la asignación de responsabilidades relacionadas con la creación de objetos, se asigna la responsabilidad de que una clase B cree un Objeto de la clase A solamente cuando:

B contiene a A.

B es una agregación (o composición) de A.

B almacena a A.

B tiene los datos de inicialización de A (datos que requiere su constructor).

Capítulo 2. Propuesta del Sistema

B usa a A.

Controlador: Asigna la responsabilidad de controlar el flujo de eventos del sistema, a clases específicas. Esto facilita la centralización de actividades (validaciones, seguridad, etc.). El controlador no realiza estas actividades, las delega en otras clases con las que mantiene un modelo de alta cohesión. Se recomienda dividir los eventos del sistema en el mayor número de controladores para poder aumentar la cohesión y disminuir el acoplamiento.

En la elaboración del presente modelo de diseño y con el objetivo de mejorar la calidad de los diagramas correspondientes a esta disciplina fueron aplicados los patrones mencionados anteriormente.

2.9.2.2 Diagrama de clases del diseño

Los diagramas de clases del diseño son una representación más concreta y detallada que los diagramas de clases del análisis, aunque también representan la parte estática del sistema conteniendo las clases y sus relaciones. Son empleados para representar las relaciones que se establecen entre las clases.

En la siguiente figura se muestra el diagrama de diseño del caso de uso Buscar_CoocurrenciaPalabras, para consultar los restantes diagramas ver anexo 7: Diagramas de clases del diseño.

Capítulo 2. Propuesta del Sistema

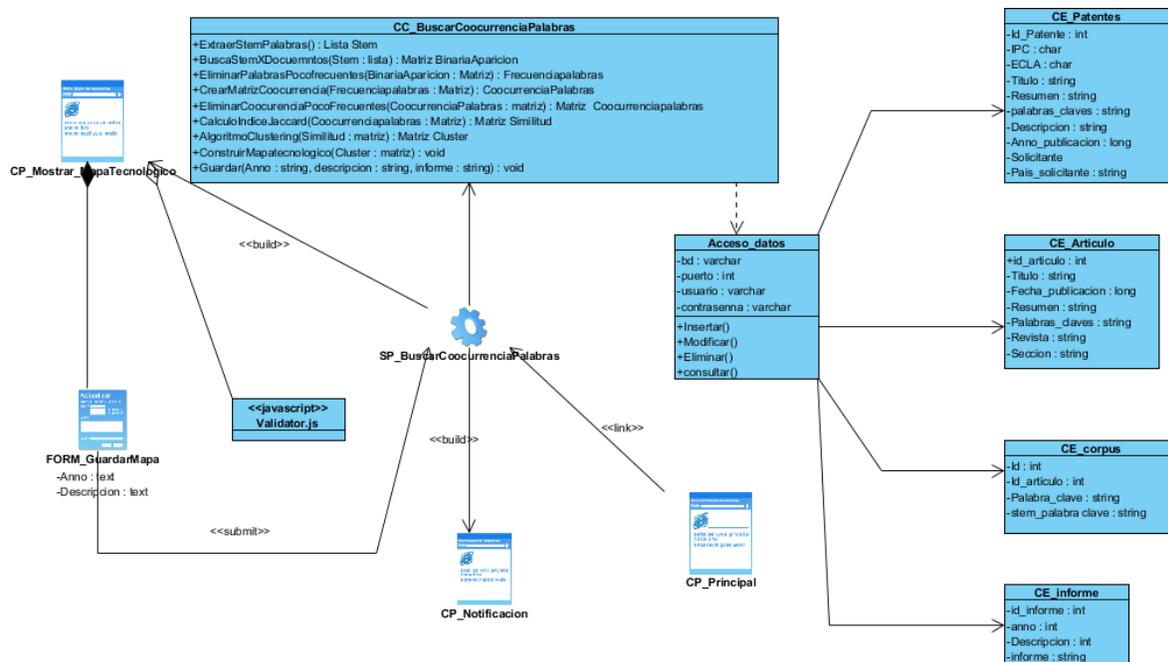


Figura 21. Diagrama de clase diseño Buscar coocurrencia de palabras.

Conclusiones Parciales

- Se propone un sistema de vigilancia tecnológica que incluye los procesos de adquisición, procesamiento y difusión de la información para ofrecer productos de valor añadido a la institución.
- El flujo de trabajo análisis y diseño permitió definir las funcionalidades previstas y las restricciones impuestas (requerimientos) así como obtener los artefactos necesarios que indican con claridad lo que se debe programar.

Capítulo 3. Validación de la propuesta

3. CAPÍTULO 3. VALIDACIÓN DE LA PROPUESTA

En el presente capítulo se realiza la validación de los requisitos y el diseño, mediante métricas para medir la calidad de los artefactos obtenidos, con el objetivo de comprobar que constituye una solución al problema planteado en la investigación.

Además de las métricas aplicadas se cuenta con una carta de aceptación por parte del grupo de calidad de la facultad donde especifica su conformidad con los requisitos. Ver Anexo 11

3.1 Métricas

En la mayoría de los desafíos técnicos, las métricas nos ayudan a entender tanto el proceso técnico que se utiliza para desarrollar un producto, como el propio producto.

“Las métricas son medidas, generalmente no absolutas pero sí objetivas, que proporcionan una forma sistemática de valorar la calidad de los elementos basándose en un conjunto de reglas bien definidas. Tienen un papel decisivo en la obtención de un producto de alta calidad, porque determinan mediante estadísticas basadas en la experiencia, el avance del software y el cumplimiento de parámetros requeridos.” [43]

“En la ingeniería del software, la medición es una disciplina relativamente joven, y no existe consenso general sobre la definición exacta de los conceptos y terminología que maneja.

A continuación se muestran los conceptos del Institute of Electrical and Electronics Engineers (IEEE) para estos términos.” [44]

Métrica: medida cuantitativa del grado en que un sistema, componente o proceso posee un atributo determinado.

Medida: valor asignado a un atributo de una entidad mediante una medición.

Medición: proceso por el cual se obtiene una medida.

Pressman propone un proceso de medición, el cual se caracteriza por cinco actividades: [43]

- **Formulación:** La obtención de medidas y métricas del software apropiadas para la representación de software en cuestión.
- **Colección:** El mecanismo empleado para acumular datos necesarios para obtener las métricas formuladas.

Capítulo 3. Validación de la propuesta

- **Análisis:** El cálculo de las métricas y la aplicación de herramientas matemáticas.
- **Interpretación:** La evaluación de los resultados de las métricas en un esfuerzo por conseguir una visión interna de la calidad de la representación.
- **Realimentación:** Recomendaciones obtenidas de la interpretación de métricas técnicas transmitidas al equipo de software.

Se realizó un estudio sobre las siguientes métricas para sistemas orientados a objetos: CK Chidamber y Kemerer [45], Lorenz y Kidd (1994) [45], Métricas R. Martin (1994) [44]; y la métrica propuesta por Alan Davis (1993) [46] que corresponde a la calidad de la Especificación de los Requisitos de Software.

Se seleccionó como técnica a utilizar en la validación de los requisitos la métrica planteada por Alan Davis, debido a que propone indicadores que coinciden con las características que un requisito debe de cumplir (ver epígrafe 2.7). Para el diseño se seleccionó la métrica de tamaño de clases (TC), propuestas por Lorenz y Kidd.

3.2 Métrica para la Calidad de la Especificación de los Requisitos.

“La Métrica para la Calidad de la Especificación de los Requisitos mide la especificidad de los requisitos, haciendo que la parte interesada pueda entenderlos de manera fácil y se puedan probar. Esta métrica incluye las siguientes características:” [46]

- **Especificidad:** ausencia de ambigüedades en la especificación, facilitando los procesos de captura y procesamiento de requerimientos. Un requisito debe tener una única interpretación.
- **Corrección:** el requerimiento debe representar algo requerido por el sistema a construirse. el grado de validación de los requerimientos mide la corrección en sus definiciones.
- **Compleción:** completitud de las definiciones.
- **Comprensión:** todo tipo de lector (clientes, usuarios, equipo de calidad desarrolladores, gestores, entre otros.) deben entender el requerimiento.
- **Consistencia interna:** no existen subconjuntos de requisitos contradictorios.
- **Consistencia externa:** ninguno de los requerimientos está en contradicción con lo expresado en documentos de nivel superior.

En el presente trabajo se seleccionaron tres revisores teniendo en cuenta los criterios siguientes:

Capítulo 3. Validación de la propuesta

- Graduado de nivel superior.
- Tres años de experiencia como mínimo en el tema de la Ingeniería de Software.
- Conocimientos acerca de los indicadores de evaluación.

Se tuvo en cuenta que si dos de los tres revisores clasifican en Correcto (C) el requisito atendiendo a los indicadores, el resultado final será correcto; de igual manera se valoran los Incorrectos (I). Para consultar el resultado emitido por los revisores ver Anexo 8.

Para aplicar la métrica, inicialmente fue necesario calcular el número total de requisitos.

$$R_t = R_f + R_{nf}$$

Donde:

R_f: Número de requisitos funcionales.

R_{nf}: Número de requisitos no funcionales.

R_t: Total de requisitos.

$$R_t = 8 + 11 = 19$$

Especificidad

Para determinar la especificidad o falta de ambigüedad de los requisitos, se sugiere una métrica basada en la consistencia de la interpretación de los revisores de cada requisito.

En la evaluación de la especificidad de los requisitos cada revisor planteó lo que había interpretado por cada requisito y se pudo apreciar que las interpretaciones coincidían, demostrando que las especificaciones de los requisitos presentan un alto grado de claridad, ya que responden a una única interpretación, no siendo necesario modificarlos.

El objetivo de esta métrica es cuantificar la especificidad o falta de ambigüedad en la definición de los requerimientos. La especificidad se calcula como:

$$Q_1 = R_{ii} / R_t$$

Siendo:

Q₁: grado de especificidad de los requisitos.

R_{ii}: Número de requisitos para los que todos los revisores tuvieron interpretaciones idénticas.

Capítulo 3. Validación de la propuesta

Cuanto más cerca de 1 esté el valor de Q_1 , mayor será la consistencia de la interpretación de los revisores para cada requisito y menor será la ambigüedad de la especificación de los requisitos.

Sustituyendo en:

$$Q_1 = 19/19$$

$$Q_1 = 1$$

Resultado: No se detectaron ambigüedades.

Corrección

El grado de validación de los requerimientos está siempre entre 0 y 1, donde el valor óptimo es el más cercano a 1; que indica un alto nivel de corrección en la definición de los requisitos. Este valor se calcula como:

$$Q_2 = R_C / (R_C + R_{NV})$$

Siendo:

Q_2 : Grado de validación de los requisitos.

R_C : Número de requisitos que se han validado como correctos.

R_{NV} : Número de requisitos que no se han validado como correctos aún.

Sustituyendo en:

$$Q_2 = 18 / (18 + 1) \text{ ver Anexo 8.1}$$

$$Q_2 = 0.94$$

Resultado: Teniendo en cuenta que el valor obtenido es cercano a 1 se plantea que los requisitos presentan un alto nivel de corrección.

Compleción

El resultado de esta métrica representa el nivel de completitud de las definiciones de los requisitos y está siempre entre 0 y 1. Grado en que los requerimientos cumplen las necesidades de los usuarios. Cuanto más próximo este el valor de 1 mayor será el nivel de completitud.

$$Q_3 = n_A / (n_A + n_B)$$

Siendo:

Capítulo 3. Validación de la propuesta

Q₃: Nivel de completitud de las definiciones de los requisitos.

n_A: Número de requisitos completos.

n_B: Número de requisitos pobremente especificados.

Sustituyendo en:

$$Q_3 = 19 / (19 + 0) \text{ ver Anexo 8.2}$$

$$Q_3 = 1$$

Resultado: Los requisitos presentan un alto grado en sus definiciones.

Comprensión

La comprensión de los requisitos marca el índice de legibilidad, no existe dificultad en su entendimiento. El resultado de esta métrica está siempre entre 0 y 1. El valor óptimo es el más cercano a 1. Se determinó a partir de la relación que se muestra a continuación.

$$Q_4 = R_{cb} / R_t$$

Siendo:

Q₄: Nivel de comprensión.

R_{bc}: Número de requisitos que todos los revisores entienden.

Sustituyendo en:

$$Q_4 = 19 / 19 \text{ ver Anexo 8.3}$$

$$Q_4 = 1$$

Resultado: Los requisitos presentan un alto grado de legibilidad.

Consistencia interna.

El resultado de esta métrica expresa que no existen subconjuntos de requisitos contradictorios (no hay conflictos), y siempre está entre 0 y 1. El valor óptimo es el más cercano a 1.

Donde:

$$Q_5 = (n_u - n_n) / n_u$$

Siendo:

n_n: Número de requisitos en conflicto con otros requisitos en la especificación.

Capítulo 3. Validación de la propuesta

n_u: Número de requisitos especificados.

Sustituyendo en:

$$Q_5 = (19 - 0) / 19 \text{ ver Anexo 8.4}$$

$$Q_5 = 1$$

Resultado: No existen requisitos en conflictos con otros.

Consistencia externa

El resultado de esta métrica está siempre entre 0 y 1. Expresa que ninguno de los requisitos está en contradicción con lo expresado en documentos de nivel superior. El valor óptimo es el más cercano a 1.

$$Q_6 = n_{ec} / R_t$$

Siendo:

n_{ec}: Número de requisitos que son consistentes con otros documentos.

Sustituyendo en:

$$Q_6 = 19 / 19 \text{ ver Anexo 8.5}$$

$$Q_6 = 1$$

Resultado: No existen requisitos en contradicción con documentos de nivel superior.

3.3 Métricas para evaluar el diseño

“Las métricas para software, como otras métricas, no son perfectas; muchos expertos argumentan que se necesita más experimentación hasta que se puedan emplear bien las métricas de diseño. Sin embargo el diseño sin medición es una alternativa inaceptable.” [47]

Tamaño de Clase (TC)

El tamaño general de una clase se puede determinar empleando métricas para saber el número total de operaciones (tanto operaciones heredadas como privadas de la instancia) que están encapsuladas dentro de la clase y encontrando el número de atributos (tanto atributos heredados como atributos privados de la Instancia) que están encapsulados en la clase. Las medidas o umbrales para los parámetros de calidad han sido una polémica a nivel mundial en el diseño de sistemas. Algunos especialistas plantean umbrales para estas métricas según como se muestra a continuación, los cuales fueron aplicados al diseño.

Capítulo 3. Validación de la propuesta

Tabla 4. Umbral

Número de Operaciones y/o Atributos.	
Métrica TC	Umbral
Pequeño	≤ 20
Medio	$> 20 \text{ y } \leq 30$
Grande	> 30

Los valores grandes de TC demuestran que una clase puede tener demasiada responsabilidad, lo cual reducirá la reutilizabilidad de la clase y complicará la implementación y la comprobación. Cuanto menor sea el valor medio para el tamaño, más probable es que las clases existentes dentro del sistema se puedan reutilizar ampliamente.

Resultados.

Al aplicar la métrica TC en las clases presentes en la capa del controlador quedaron recogidos los siguientes datos:

Tabla 5. Resultados del umbral

Clases	Atributos	Operaciones	Umbral
Gestionar_fuente	0	14	Pequeño
Gestionar_autor	0	5	Pequeño
Normalizar_palabras_claves	0	8	Pequeño
Contar_palabras	0	5	Pequeño
Contar_palabrasClaves_año	0	7	Pequeño
Contar_autores_pCaño	0	8	Pequeño
Buscar_coocurrencia_palabra	0	9	Pequeño
Modelar_relaciones_sociales	0	9	Pequeño

Todas las clases que conforman esta capa están dentro de la categoría de pequeño, lo que demuestra que el diseño del sistema no es complejo, quedando demostrada la calidad del diseño.

Conclusiones parciales

La definición y aplicación de una estrategia de prueba permitió la validación de la especificación de los requisitos y el diseño, lo que propiciará una mayor confiabilidad al implementar el sistema.

CONCLUSIONES GENERALES

- En las actividades de vigilancia se utiliza de manera creciente las aportaciones de la bibliometría y la cienciometría para realizar el tratamiento de grandes volúmenes de información con la ayuda de herramientas informáticas.
- Los sistemas de vigilancia son capaces de identificar empresas, tendencias y tecnologías a partir de métodos estadísticos, proporcionando elementos para orientar decisiones estratégicas.
- Se propone un sistema de vigilancia tecnológica que incluye los procesos de adquisición, procesamiento y difusión de la información para ofrecer productos de valor añadido a la institución.
- La validación de la propuesta elaborada asintió la aptitud del Análisis y Diseño como guía para estructurar el proceso de desarrollo del sistema.

RECOMENDACIONES

- Implementar a partir de la propuesta realizada una herramienta de vigilancia tecnológica que permita la adecuada gestión de la información.
- Fomentar el uso de la vigilancia tecnológica con el empleo de herramientas para el descubrimiento de conocimiento, que contribuya a anticiparse a los cambios y la toma de decisiones de la institución.
- Automatizar el proceso de recuperación de las fuentes de información.
- Explotar el análisis de las patentes como una amplia fuente de información tecnológica.
- Capacitar a los investigadores del grupo GIC sobre los conceptos, métodos y herramientas relacionados con la vigilancia tecnológica, para que estén en condiciones de asimilar este tipo de procesos y los implementen de manera adecuada en el Centro.

REFERENCIAS BIBLIOGRÁFICAS

- [1] CDE. CDE tu mejor aliado. [Online]. http://www.cde.es/es/inteligencia_competitiva/implantaciones_de_sistemas_de_vt_ic/en_empresas/
- [2] IDICT VILLA CLARA. IDICT VILLA CLARA. [Online]. <http://www.idict.villaclara.cu/productos-servicios/cartera-productos-servicios/consultorias/vigilancia-tecnologica>
- [3] Tamara León Aliz. Biblioteca Virtual de las Ciencias en Cuba. [Online]. <http://www.bibliociencias.cu/gsd/collect/eventos/archives/HASH6393.dir/doc.pdf>
- [4] Yoel Abreu Marta Infante, Mercedes Delgado, and Olga Infante, "Minería tecnológica para el análisis de oportunidades de publicaciones en la universidad," Instituto Superior Politécnico "José Antonio Echeverría",.
- [5] PROYECTO ESTRATEGICO, "PROYECTO ESTRATEGICO 2008-2012," Universidad de las Ciencias Informáticas., La Habana-Cuba.,.
- [6] Centro FORTES. Redmine del Centro FORTES. [Online]. <http://gespro.fortes.prod.uci.cu/>
- [7] González Limas, *Seminario Nacional de la OMPI sobre Propiedad Industrial, Invenciones e Información Tecnológica*. Madrid: Fundacion Cervantes, 2002.
- [8] Rouach, "Vigilancia e Inteligencia Competitiva: Fundamentos e implicaciones," *madrimsd*, vol. 17, junio-julio 2003.
- [9] Fernando Palop y José M. Vicente, "VIGILANCIA TECNOLÓGICA INTELIGENCIA E COMPETITIVA SU POTENCIAL PARA LA EMPRESA ESPAÑOLA," Febrero, 1999.
- [10] Odalis Bouza Betancourt, "Desarrollo del ambito informacional desde la perspectiva de la sistematizacion de la vigilancia y tecnologica(VCT) en organizaciones," Universidad de la Habana, Tesis doctoral 2010.
- [11] AV Strategos. AV Strategos. [Online]. http://portal.strategos.es/index.php?option=com_content&view=article&id=79&Itemid=192
- [12] RAMON MASPONS BOSCH PERE ESCORSA CASTELLS, "LA VIGILANCIA TECNOLÓGICA, UN REQUISITO INDISPENSABLE PARA LA INNOVACION," EOIAMERICA.,.
- [13] Martha Beatriz Infante Abreu, Olga Infante Pérez, Yoel Abreu Lee, Beatriz M. García Delgado Mercedes Delgado Fernández, "Vigilancia tecnológica como factor clave para el éxito en la I+D+i: aplicación en el ámbito universitario," Facultad de Ingeniería Industrial, Instituto Superior Politécnico "José Antonio Echeverría" CUJAE , Habana.,.
- [14] Martha Beatriz Infante Abreu, Olga Infante Pérez, Yoel Abreu Lee, Beatriz M. García Delgado Mercedes Delgado Fernández, "Minería tecnológica para el análisis de oportunidades de publicaciones en la universidad.,," *revista.cnic.edu.cu*, p. 3.
- [15] Linnet Ivet Bello Leyva, "Propuesta de proceso de Vigilancia Tecnológica para la implementación de un Observatorio de Tecnología Educativa en el Centro de Tecnologías para la Formación," UCI, Tesis de grado 2010.
- [16] "Estudios de vigilancia tecnológica aplicados a cadenas productivas del sector agropecuario colombiano," BOGOTÁ., 2008.
- [17] Francisco Téllez, "Redes Formales e Informales de Producción y Diseminación de Información," 7 y 8 de noviembre del 2005.
- [18] Pilar Lázaro, "La inteligencia competitiva, factor clave para la toma de decisiones estratégicas en las organizaciones.,," in *La inteligencia competitiva, factor clave para la toma de decisiones estratégicas en las organizaciones*. Madrid : Plan Regional de Ciencia y Tecnología de la Comunidad de Madrid, 2007-2008.
- [19] MSc, Fabio A. González O. PhD. Víctor A. Bucheli G., "Herramienta informática para Vigilancia Tecnológica-VIGITECH.,," Universidad Nacional de Colombia, sede Bogotá Departamento de Ingeniería de Sistemas e Industrial , 2007.
- [20] Gema Cossío Cárdenas, "Las herramientas de software para el análisis cuantitativo de información. Estudio preliminar de aplicaciones desarrolladas en Cuba," La Habana, 2009.

Referencias bibliográficas

- [21] (2011) Buenas tareas.Herramientas Case. [Online]. <http://www.buenastareas.com/ensayos/Herramientas-Case/3184054.html>
- [22] Maritza Silveira Orozco and Katia Pardo Duarte, "Análisis y Diseño del Sistema de Gestión Tecnológica de la Facultad Regional de Granma", Universidad de las Ciencias Informáticas(UCI), La Habana-Cuba, 2010.
- [23] GRUPO SOLUCIONES GSI NNOVA. [Online]. <http://www.rational.com.ar/herramientas/roseenterprise.html>
- [24] CENTERS FOR MEDICARE AND MEDICAID SERVICES., "SELECTING A DEVELOPMENT APPROACH.", 2008.
- [25] Universidad de San Martín de Porres. (2012) Comparando RUP Y XP. [Online]. <http://www.usmp.edu.pe/>
- [26] Dr.Homero R Saltalamacchia, *Del proyecto al análisis: aportes a la investigación cualitativa.*, 3rd ed.
- [27] Kryscia Daviana Ramírez Benavides, "Stemming – Lematización",.
- [28] Xavier Garcia Cuerda. Introducción a los Sistemas de Gestión de Contenidos (CMS) de código abierto. [Online]. <http://mosaic.uoc.edu/2004/11/29/introduccion-a-los-sistemas-de-gestion-de-contenidos-cms-de-codigo-abierto/>
- [29] Cliente y Servidor. [Online]. http://docente.ucol.mx/rcarlos/public_html/cliente-servidor.htm
- [30] (2011.) Buenas Tareas.Lenguaje Html. [Online]. <http://www.buenastareas.com/ensayos/Lenguaje-Html/2757842.html>
- [31] LibrosWeb.es.Introducción a JavaScript. [Online]. <http://www.librosweb.es/javascript/>
- [32] LibrosWeb. LibrosWeb.es.Introducción a CSS. [Online]. <http://www.librosweb.es/css/>
- [33] Maestros del Web. Maestros del Web. [Online]. <http://www.maestrosdelweb.com/editorial/phpintro/>
- [34] Ian Gilfillan, *La Biblia de Mysql.*
- [35] Aliuska Tamayo Cordovés, "Intranet del Ministerio de la Industria Alimentaria," UCI, La Habana, 2011.
- [36] Mauricio Arévalo M, "Ventajas y Desventajas de PostgreSQL",.
- [37] Aplicaciones Empresariales. (2012) AplicacionesEmpresariales.com. [Online]. <http://www.aplicacionesempresariales.com/ventajas-y-desventajas-de-postgresql.html>
- [38] NETCRAFT. (2011) NETCRAFT. [Online]. <http://news.netcraft.com/archives/2011/10/06/october-2011-web-server-survey.html>
- [39] Michel Arias Chavez, *LA INGENIERIA DE REQUERIMIENTOS Y SU IMPORTANCIA EN EL DESARROLLO DE PROYECTOS DE SOFTWARE.*, 2006, vol. 6.
- [40] "Maestría en Ingeniería de Sistemas con Mención en Gestión En Tecnología de la Información," Limas, 2007.
- [41] EcuRed. (2012) [Online]. http://www.ecured.cu/index.php/Patrones_de_Casos_de_Uso
- [42] Ivar Jacobson, Grady Booch, and James Rumbaugh, *El Proceso Unificado de Desarrollo de Software.* Madrid: Pearson Educación, S.A, 2000.
- [43] Roger Presman., "Métricas del Software.,".
- [44] Ana María García Sánchez, "Evaluación de métricas de calidad del software sobre un programa Java.," Dpto. de Sistemas Informáticos y Computación, Proyecto de Fin de Máster en Programación y Tecnología Software 2010.
- [45] Ed Berard, "MÉTRICAS TÉCNICAS PARA SISTEMAS ORIENTADOS A OBJETOS,".
- [46] Alan Davis, "La Especificación de Requerimientos de Software," in *INGENIERIA DE SOFTWARE I*, Guía 2, Ed., 1993.
- [47] *Métricas en el desarrollo del Software*, 4th ed.

BIBLIOGRAFÍA

- [1] CDE. CDE tu mejor aliado. [Online]. http://www.cde.es/es/inteligencia_competitiva/implantaciones_de_sistemas_de_vt_ic/en_empresas/
- [2] IDICT VILLA CLARA. IDICT VILLA CLARA. [Online]. <http://www.idict.villaclara.cu/productos-servicios/cartera-productos-servicios/consultorias/vigilancia-tecnologica>
- [3] Tamara León Aliz. Biblioteca Virtual de las Ciencias en Cuba. [Online]. <http://www.bibliociencias.cu/gsd/collect/eventos/archives/HASH6393.dir/doc.pdf>
- [4] Yoel Abreu Marta Infante, Mercedes Delgado, and Olga Infante, "Minería tecnológica para el análisis de oportunidades de publicaciones en la universidad," Instituto Superior Politécnico "José Antonio Echeverría",.
- [5] PROYECTO ESTRATEGICO, "PROYECTO ESTRATEGICO 2008-2012," Universidad de las Ciencias Informáticas., La Habana-Cuba.,.
- [6] Centro FORTES. Redmine del Centro FORTES. [Online]. <http://gespro.fortes.prod.uci.cu/>
- [7] González Limas, *Seminario Nacional de la OMPI sobre Propiedad Industrial, Invenciones e Información Tecnológica*. Madrid: Fundacion Cervantes, 2002.
- [8] Rouach, "Vigilancia e Inteligencia Competitiva: Fundamentos e implicaciones," *madrimasd*, vol. 17, junio-julio 2003.
- [9] Fernando Palop y José M. Vicente, "VIGILANCIA TECNOLÓGICA INTELIGENCIA E COMPETITIVA SU POTENCIAL PARA LA EMPRESA ESPAÑOLA," Febrero, 1999.
- [10] Odalis Bouza Betancourt, "Desarrollo del ambito informacional desde la perspectiva de la sistematizacion de la vigilancia y tecnologica(VCT) en organizaciones," Universidad de la Habana, Tesis doctoral 2010.
- [11] AV Strategos. AV Strategos. [Online]. http://portal.strategos.es/index.php?option=com_content&view=article&id=79&Itemid=192
- [12] RAMON MASPONS BOSCH PERE ESCORSA CASTELLS, "LA VIGILANCIA TECNOLÓGICA, UN REQUISITO INDISPENSABLE PARA LA INNOVACION," EOIAMERICA.,.
- [13] Martha Beatriz Infante Abreu, Olga Infante Pérez, Yoel Abreu Lee, Beatriz M. García Delgado Mercedes Delgado Fernández, "Vigilancia tecnológica como factor clave para el éxito en la I+D+i: aplicación en el ámbito universitario," Facultad de Ingeniería Industrial, Instituto Superior Politécnico "José Antonio Echeverría" CUJAE , Habana.,.
- [14] Martha Beatriz Infante Abreu, Olga Infante Pérez, Yoel Abreu Lee, Beatriz M. García Delgado Mercedes Delgado Fernández, "Minería tecnológica para el análisis de oportunidades de publicaciones en la universidad.," *revista.cnic.edu.cu*, p. 3.
- [15] Linnet Ivet Bello Leyva, "Propuesta de proceso de Vigilancia Tecnológica para la implementación de un Observatorio de Tecnología Educativa en el Centro de Tecnologías para la Formación," UCI, Tesis de grado 2010.
- [16] "Estudios de vigilancia tecnológica aplicados a cadenas productivas del sector agropecuario colombiano," BOGOTÁ., 2008.
- [17] Francisco Téllez, "Redes Formales e Informales de Producción y Diseminación de Información," 7 y 8 de noviembre del 2005.
- [18] Pilar Lázaro, "La inteligencia competitiva, factor clave para la toma de decisiones estratégicas en las organizaciones.," in *La inteligencia competitiva, factor clave para la toma de decisiones estratégicas en las organizaciones*. Madrid : Plan Regional de Ciencia y Tecnología de la Comunidad de Madrid, 2007-2008.
- [19] MSc, Fabio A. González O. PhD. Víctor A. Bucheli G., "Herramienta informática para Vigilancia Tecnológica-VIGITECH.," Universidad Nacional de Colombia, sede Bogotá Departamento de Ingeniería de Sistemas e Industrial , 2007.

- [20] Gema Cossío Cárdenas, "Las herramientas de software para el análisis cuantitativo de información. Estudio preliminar de aplicaciones desarrolladas en Cuba," La Habana, 2009.
- [21] (2011) Buenas tareas.Herramientas Case. [Online]. <http://www.buenastareas.com/ensayos/Herramientas-Case/3184054.html>
- [22] Maritza Silveira Orozco and Katia Pardo Duarte, ""Análisis y Diseño del Sistema de Gestión Tecnológica de la Facultad Regional de Granma",," Universidad de las Ciencias Informáticas(UCI)., La Habana-Cuba, 2010.
- [23] GRUPO SOLUCIONES GSI NNOVA. [Online]. <http://www.rational.com.ar/herramientas/roseenterprise.html>
- [24] CENTERS FOR MEDICARE AND MEDICAID SERVICES., "SELECTING A DEVELOPMENT APPROACH.," 2008.
- [25] Universidad de San Martín de Porres. (2012) Comparando RUP Y XP. [Online]. <http://www.usmp.edu.pe/>
- [26] Dr.Homero R Saltalamacchia, *Del proyecto al análisis: aportes a la investigación cualitativa.*, 3rd ed.
- [27] Kryscia Daviana Ramírez Benavides, "Stemming – Lematización,".
- [28] Xavier Garcia Cuerda. Introducción a los Sistemas de Gestión de Contenidos (CMS) de código abierto. [Online]. <http://mosaic.uoc.edu/2004/11/29/introduccion-a-los-sistemas-de-gestion-de-contenidos-cms-de-codigo-abierto/>
- [29] Cliente y Servidor. [Online]. http://docente.ucol.mx/rcarlos/public_html/cliente-servidor.htm
- [30] (2011.) Buenas Tareas.Lenguaje Html. [Online]. <http://www.buenastareas.com/ensayos/Lenguaje-Html/2757842.html>
- [31] LibrosWeb.es.Introducción a JavaScript. [Online]. <http://www.librosweb.es/javascript/>
- [32] LibrosWeb. LibrosWeb.es.Introducción a CSS. [Online]. <http://www.librosweb.es/css/>
- [33] Maestros del Web. Maestros del Web. [Online]. <http://www.maestrosdelweb.com/editorial/phpintro/>
- [34] Ian Gilfillan, *La Biblia de Mysql.*
- [35] Aliuska Tamayo Cordovés, "Intranet del Ministerio de la Industria Alimentaria," UCI, La Habana, 2011.
- [36] Mauricio Arévalo M, "Ventajas y Desventajas de PostgreSQL,".
- [37] Aplicaciones Empresariales. (2012) AplicacionesEmpresariales.com. [Online]. <http://www.aplicacionesempresariales.com/ventajas-y-desventajas-de-postgresql.html>
- [38] NETCRAFT. (2011) NETCRAFT. [Online]. <http://news.netcraft.com/archives/2011/10/06/october-2011-web-server-survey.html>
- [39] Michel Arias Chavez, *LA INGENIERIA DE REQUERIMIENTOS Y SU IMPORTANCIA EN EL DESARROLLO DE PROYECTOS DE SOFTWARE.*, 2006, vol. 6.
- [40] "Maestría en Ingeniería de Sistemas con Mención en Gestión En Tecnología de la Información," Limas, 2007.
- [41] EcuRed. (2012) [Online]. http://www.ecured.cu/index.php/Patrones_de_Casos_de_Uso
- [42] Ivar Jacobson, Grady Booch, and James Rumbaugh, *El Proceso Unificado de Desarrollo de Software.* Madrid: Pearson Educación, S.A, 2000.
- [43] Roger Presman., "Métricas del Software.,".
- [44] Ana María García Sánchez, "Evaluación de métricas de calidad del software sobre un programa Java.," Dpto. de Sistemas Informáticos y Computación, Proyecto de Fin de Máster en Programación y Tecnología Software 2010.
- [45] Ed Berard, "MÉTRICAS TÉCNICAS PARA SISTEMAS ORIENTADOS A OBJETOS,".
- [46] Alan Davis, "La Especificación de Requerimientos de Software," in *INGENIERIA DE SOFTWARE I*, Guia 2, Ed., 1993.
- [47] *Métricas en el desarrollo del Software*, 4th ed.

- [48] Richard Chiroque-Solano y Paola Padilla-Santoyo, "Análisis de coautoría en la revista Biblios: Una aproximación desde Google Scholar ," Escuela de Bibliotecología y Ciencias de la Información. Universidad Nacional Mayor de San Marcos ,.
- [49] Pere Escorsa Castells, "De la vigilancia tecnológica a la inteligencia competitiva en las empresas.," 2001.
- [50] Ángel F. Zazo, Emilio Rodríguez Vázquez de Aldana, José Luis Alonso Berrocal Carlos G. Figuerola, "La Recuperación de Información en español y la normalización de términos.," *Revista Iberoamericana de Inteligencia Artificial.* , vol. VIII, no. 22, 2004.
- [51] Sergio Larreina García, Sandra Hernando, and David Grisaleña, "La Evolución de la Inteligencia Competitiva," *PUZZLE.Revista Hispana de la Inteligencia Competitiva:Un estudio de las herramientas cuantitativas.*, vol. V, no. 020, 2006.
- [52] Boston College Steve Borgatti, "Conceptos Básicos de Redes Sociales," Cancún, 2003.
- [53] Alejandro Velázquez Álvarez and Norman Aguilar Gallegos, "Manual Introductorio al Análisis de Redes Sociales. ," 2005.
- [54] Cristian Paolo Mejía Olivares, "Análisis de Redes Sociales a Gran Escala," CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL DEPARTAMENTO DE COMPUTACIÓN, Tesis de Maestría. 2010.
- [55] COTEC, *FUNDACIÓN COTEC P ARA LA INNOVACIÓN TECNOLÓGICA.LOS INFORMES TECNOLÓGICOS DE PATENTES.*, I ed., 2007.
- [56] Francisca Losavio, and Alfredo Matteo Ledis Chirinos, "IDENTIFYING QUALITY-BASED REQUIREMENTS,".

GLOSARIO DE TÉRMINOS

Agrupamiento en inglés clustering: proceso de dividir un conjunto de datos en grupos mutuamente excluyentes de tal manera que cada miembro de un grupo esté lo “más cercano” posible a otro, y grupos diferentes estén lo “más lejos” posible uno del otro, donde la distancia está medida con respecto a todas las variables disponibles.

Análisis morfológico: consiste en determinar la forma, clase o categoría gramatical de cada palabra.

Colecciones documentales: es la reunión artificial de documentos acumulados sobre la base de algunas características comunes, sin tener en cuenta la procedencia de dichos documentos.

GPL: es una licencia creada por la Free Software Foundation en 1989 (la primera versión, escrita por Richard Stallman), y está orientada principalmente a proteger la libre distribución, modificación y uso de software.

Grafos: es un conjunto, no vacío, de objetos llamados vértices (o nodos) y una selección de pares de vértices, llamados aristas (*edges* en inglés) que pueden ser orientados (dirigidos) o no.

Lematización: es un proceso de normalización lingüística en el cual las diferentes formas que puede adoptar una palabra son reducidos a una única forma común, a la cual se denomina stem, lema o raíz de la palabra.

Nichos de mercado: es una oportunidad que brinda la economía para desarrollar una cierta actividad comercial o productiva con elevadas posibilidades de éxito ante las condiciones del mercado.

Procesamiento de información: es la aplicación sistemática de una serie de operaciones sobre un conjunto de datos, generalmente por medio de máquinas, para explotar la información que estos datos representan.

Recuperación de información: es la ciencia de la búsqueda de información en documentos electrónicos y cualquier tipo de colección documental digital, encargada de la búsqueda dentro de éstos mismos, búsqueda de metadatos que describan documentos, o también la búsqueda en bases de datos relacionales, ya sea a través de internet, intranet, y como objetivo realiza la recuperación en textos, imágenes, sonido o datos de otras

Glosario de términos

características, de manera pertinente y relevante.

Relaciones cognitivas: se refiere a una representación conceptual de los objetos.

Stem: es la porción de la palabra después de eliminar sus afijos.

Teoría de grafos: estudia las propiedades de los grafos.

Tesauros: Instrumento constituido por un repertorio alfabético de términos normalizados de un vocabulario generalmente especializado con el fin de analizar o indicar la información relativa a los documentos, series, fondos, entre otros.

ANEXOS

Anexo 1. Conceptos abordados por teóricos de vigilancia tecnológica.

Tabla 6. Conceptos de vigilancia tecnológica.

Autores	Definiciones
Jakobiak (1992)	<i>"La vigilancia tecnológica consiste en la observación y en el análisis del entorno científico, tecnológico y de los impactos económicos presentes y futuros para identificar las amenazas y oportunidades de desarrollo". [8]</i>
Lesca (1994)	<i>"La vigilancia tecnológica incluye los esfuerzos que la empresa dedica, los medios de que se dota y las disposiciones que toma con el objetivo de conocer todas las evoluciones y novedades que se producen en los dominios de las técnicas que le conciernen actualmente o son susceptibles de afectarle en el futuro". [8]</i>
Werner & Degoul (1994)	<i>"Es el medio de hacer emerger los elementos estratégicos para la empresa de entre la masa de información disponible" [11]</i>
Martinet y Martí, (1995)	<i>"La vigilancia tecnológica permite a la empresa determinar los sectores de donde vendrán las mayores innovaciones tanto para los procesos como para los productos que tienen incidencia en la empresa". [8]</i>
Escorsa y Maspons 2001	<i>La vigilancia tecnológica "consiste en realizar de forma sistemática la captura, el análisis, la difusión y la explotación de las informaciones técnicas útiles para la supervivencia y el crecimiento de la empresa. La vigilancia debe alertar sobre cualquier innovación científica o técnica susceptible de crear oportunidades o amenazas". [12]</i>
Morcillo, 2003	<i>Proceso de obtención, análisis, validación y difusión de información de valor estratégico sobre la organización, que se transmite a los responsables de la organización para la</i>

	<i>toma de decisión en el momento adecuado, partiendo fundamentalmente de la identificación del problema a analizar, determinando los objetivos de Vigilancia, que conllevan a la determinación de las fuentes de información para poder captarla. [48]</i>
Wheelwright	<i>“Está constituida por el conjunto de técnicas que permiten organizar de manera sistemática la acumulación, el análisis, la difusión y la explotación de las informaciones técnicas útiles para la supervivencia y crecimiento de la empresa. Tiene la misión de alertar a los responsables de la empresa de toda innovación científica o técnica susceptible de modificar su entorno” [11]</i>

Anexo 2. Comparación de las herramientas Matheo software.

Comparativa entre las tres herramientas del Matheo: Analyzer, Patent y Web

MATHEO ANALIZER	MATHEO PATENT	MATHEO WEB
<ul style="list-style-type: none"> · Minería de Datos textuales. · Base datos estructurada (patentes, artículos...). · Necesita etapa de pretratamiento (obtención de corpus e importación). · Tratamiento semiautomático de la información, lo que le concede mayor flexibilidad. · Uso de métricas estadísticas. · Visualización: histogramas, grafos y matrices. · El analista debe organizar y editar el informe. 	<ul style="list-style-type: none"> · Minería de datos textuales. · Base de datos de patentes de USPTO y EPO (limitación). · No necesita de pretratamiento de la información. · Tratamiento automático de las patentes recuperadas. · Uso de métricas estadísticas. · Visualización: histogramas, grafos y matrices. · Publicación automática de informes y reportes. 	<ul style="list-style-type: none"> · Minería de la Web. · Diversas fuentes en Internet. · No necesita de pretratamiento de la información. · Uso de métricas estadísticas. · Visualización: histogramas, grafos y matrices. · Publicación automática de informes y reportes.

Anexo 3. Descripciones de los casos de uso.

Tabla 7. Descripción del caso de uso Gestionar Fuentes.

Caso de Uso:	Gestionar fuentes.
Actores:	Captador.
Resumen:	El caso de uso comienza cuando el captador va a insertar, modificar o eliminar los metadatos de las fuentes de información.
Precondiciones:	Los metadatos de la información deben estar

	disponibles.
Referencias:	RF-1.1, RF-1.2, RF-1.3.
Prioridad:	Media.
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
1. Selecciona gestionar fuente.	Muestra las siguientes opciones: -Insertar fuente. -Modificar fuente. -Eliminar fuente.
Flujo Normal de Eventos	
Sección "Insertar Fuentes"	
Acción del Actor	Respuesta del Sistema
1. Selecciona insertar fuente.	1.1. Muestra los tipos de fuente que puede insertar que pueden ser: -Patentes. -Artículos. -Tesis. -Libros. -Ferias.
2. Selecciona el tipo de fuente a insertar.	2.1. Muestra los campos correspondientes a la fuente.
3. Inserta los datos correspondientes y presiona guardar.	3.1. Muestra notificación de la correcta inserción.
Flujo Normal de Eventos	
Sección "Modificar fuente"	
Acción del Actor	Respuesta del Sistema
4. Selecciona modificar fuente.	4.1. Muestra los tipos de fuente existentes que pueden ser: -Patentes. -Artículos. -Tesis. -Libros. -Ferias.
5. Selecciona tipo de fuente a modificar.	5.1. Muestra los campos de búsqueda de la fuente seleccionada que pueden ser por: -Título. -Autor. -Palabras claves. -Código de patentes.
6. Entra el criterio de búsqueda.	6.1. Muestra los resultados de búsqueda.
7. Selecciona la fuente a modificar.	7.1. Muestra los campos a modificar.
8. Modifica los datos y presiona guardar.	8.1. Muestra notificación de la correcta modificación.
Flujo Normal de Eventos	
Sección "Eliminar fuente"	
Acción del Actor	Respuesta del Sistema
9. Selecciona eliminar fuente.	9.1. Muestra los tipos de fuente existentes que pueden ser: -Patentes. -Artículos. -Tesis. -Libros. -Ferias.
10. Selecciona el tipo de fuente a	10.1. Muestra los campos de búsqueda de la fuente

eliminar.	seleccionada que pueden ser por: -Título. -Autor. -Palabras claves. -Código de patentes.
11. Entra el criterio de búsqueda.	11.1. Muestra los resultados de búsqueda.
12. Selecciona la o las fuente/s a eliminar y oprime aceptar.	12.1. Muestra notificación de la eliminación.
Flujos Alternos	
Campos en blanco	
2.a.1 No selecciona tipo de fuente a insertar.	2.a.2 Muestra pantalla con tipos de fuentes existentes.
5.a.1 No selecciona tipo de fuente a modificar.	5.a.2 Muestra pantalla con tipos de fuentes existentes.
10 a.1 No selecciona tipo de fuente a eliminar.	10.a.2 Muestra pantalla con tipos de fuentes existentes.
6.a.1 No inserta el criterio de búsqueda para modificar fuente y presiona aceptar.	6.a.2 Muestra pantalla con mensaje de error Muestra pantalla con los campos de búsqueda.
11.a.1 No inserta el criterio de búsqueda para eliminar fuente y presiona aceptar.	11.a.2 Muestra pantalla con mensaje de error Muestra pantalla con los campos de búsqueda.
7. No selecciona la fuente a modificar y presiona aceptar.	7.a.2 Muestra pantalla con mensaje de error Muestra pantalla con lista de fuentes a modificar.
12. No selecciona la o las fuente/s a eliminar y presiona aceptar.	12.a.2 Muestra pantalla con mensaje de error Muestra pantalla con lista de fuentes a eliminar.
El captador selecciona la opción Cancelar	
3.a.1 Entra los datos y cancela el proceso de inserción de fuentes.	3.a.2 Muestra notificación de la cancelación. 3.a.3 Muestra pantalla principal.
8.a.1 Entra los datos y cancela el proceso de modificación de la fuentes.	8.a.2 Muestra notificación. 8.a.3 Muestra pantalla principal.
12.a.1 Selecciona la fuente y cancela proceso de eliminación.	12.a.2 Muestra notificación. 12.a.3 Muestra pantalla principal.
Poscondiciones	Quedan gestionadas las fuentes de información.

Tabla 8. Descripción del caso de uso Gestionar Autor.

Caso de Uso:	Gestionar autor.
Actores:	Captador.
Resumen:	El caso de uso comienza cuando el captador va a insertar, modificar o eliminar los metadatos de un autor.
Precondiciones:	Los metadatos del autor deben estar disponibles.
Referencias:	RF-2.1, RF-2.2, RF-2.3.
Prioridad:	Media.
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema.
1. Selecciona gestionar autor	Muestra las siguientes opciones: -Insertar autor. -Modificar autor. -Eliminar autor.
Flujo Normal de Eventos	
Sección "Insertar autor"	
Acción del Actor	Respuesta del Sistema.

1. Selecciona insertar autor.	1.1. Muestra campos.
2. Inserta los datos correspondientes y presiona guardar.	2.1. Muestra notificación de la correcta inserción.
Flujo Normal de Eventos	
Sección “Modificar autor”	
Acción del Actor	Respuesta del Sistema
3. Selecciona modificar autor.	3.1. Muestra campos de búsqueda.
4. Entra el criterio de búsqueda.	4.1. Muestra los resultados de búsqueda.
5. Selecciona autor a modificar.	5.1. Extrae autor de la tabla. Muestra los campos a modificar.
6. Modifica los datos y presiona guardar.	6.1. Muestra notificación de la correcta modificación.
Flujo Normal de Eventos	
Sección “Eliminar autor”	
Acción del Actor	Respuesta del Sistema
9. Selecciona eliminar autor.	9.1 Muestra campos de búsqueda.
10. Entra el criterio de búsqueda.	10.1. Muestra resultado.
11. Selecciona el autor a eliminar y oprime aceptar.	11.1. Muestra notificación de la eliminación.
Flujos Alternos	
Campos en blanco	
2.a.1 No inserta los datos del autor y presiona aceptar.	2.a.2 Muestra pantalla con mensaje de error Muestra pantalla con los campos.
4.a.1 No inserta el criterio de búsqueda para modificar autor y presiona aceptar.	4.a.2 Muestra pantalla con mensaje de error Muestra pantalla con los campos de búsqueda.
6.a.1 No modifica los datos del autor y presiona aceptar.	6.a.2 Muestra pantalla con mensaje de error Muestra pantalla con los campos.
5.a.1 No selecciona autor a modificar.	5.a.2 Muestra pantalla con mensaje de error Muestra pantalla con autor a modificar.
10.a.1 No inserta el criterio de búsqueda para eliminar autor y presiona aceptar.	10.a.2 Muestra pantalla con mensaje de error Muestra pantalla con los campos de búsqueda.
11. No selecciona el autor a eliminar y presiona aceptar.	11.a.2 Muestra pantalla con mensaje de error Muestra pantalla con autor a eliminar.
El captador selecciona la opción Cancelar	
2.a.1 Inserta los datos del autor y presiona cancelar.	2.a.2 Muestra notificación de la cancelación. 2.a.3 Muestra pantalla principal.
4.a.1 Inserta el criterio de búsqueda para modificar autor y presiona cancelar.	4.a.2 Muestra notificación de la cancelación. 4.a.3 Muestra pantalla principal.
6.a.1 Modifica los datos del autor y presiona cancelar.	6.a.2 Muestra notificación. 6.a.3 Muestra pantalla de resultados.
5.a.1 Selecciona autor a modificar y presiona cancelar.	5.a.2 Muestra notificación. 5.a.3 Muestra pantalla de resultados.
10.a.1 Inserta el criterio de búsqueda para eliminar autor y presiona cancelar.	10.a.2 Muestra notificación. 10.a.3 Muestra pantalla principal.
11.a.1 Selecciona el autor a eliminar y presiona cancelar.	11.a.2 Muestra notificación. 11.a.3 Muestra pantalla de resultados.
Poscondiciones	Quedan gestionados los autores

Tabla 9. Descripción del caso de uso Normalizar palabras claves.

Caso de Uso:	Normalizar palabras claves.
Actores:	Experto.

Resumen:	El caso de uso comienza cuando el actor selecciona la funcionalidad normalizar campos.
Precondiciones:	El campo palabras claves debe estar lleno de las fuentes patentes y artículos científicos.
Referencias:	RF-3.
Prioridad:	Alta.
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
1 Selecciona normalizar palabras claves.	1.1 Extrae palabras claves.
	2. Guarda las palabras claves compuestas en una lista.
	2.1 Divide las palabras claves compuestas y guarda en una lista las palabras claves simples.
	3. Crea matriz de correlación de coeficiente de DICE de palabras claves simples.
	3.1 Halla digrama (lista de palabras simples).
	4. Calcula coeficiente de DICE(Lista de Digramas, lista de palabras simples, matriz de correlación de palabras simples) (Ver anexo 9).
	4.1 Determina stem (matriz de correlación de palabras simples, lista de digramas, listas de palabras simples).
	5. Genera stem de palabras compuestas (lista de palabras simples, lista de palabras clave).
	5.1 Construye tabla de resultados (id_articulo, palabra clave, stem de la palabra clave).
	6. Guarda la tabla.
	Muestra la tabla.
	6.1 Muestra notificación y termina el caso de uso.
Poscondiciones	Queda construido el corpus con los siguientes datos, id, id del artículo o código de la patente, palabras claves y la raíz de la palabra clave.

Tabla 10. Descripción del caso de uso Contar palabras.

Caso de Uso:	Contar palabras.
Actores:	Experto.
Resumen:	El caso de uso comienza cuando el actor selecciona la funcionalidad frecuencia de palabras claves.
Precondiciones:	Estar normalizadas las palabras claves.
Referencias:	RF-4.
Prioridad:	Alta.
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
1 Selecciona la opción frecuencia de palabras claves.	2 Extrae raíz de las palabras claves.
	2.1 Crea matriz de mxn donde m son las palabras claves y n los documentos.
	Busca palabras claves en los documentos usando la raíz.
	3 Llena matriz de palabras claves x documentos.
	3.1 Crea tablas de frecuencias absolutas.
	4 Cuenta ocurrencia de palabras claves en los documento.
	4.1 Construye histogramas de las tablas de frecuencias.(ver

	anexo 10)
5 Selecciona la opción de guardar.	5.1 Muestra pantalla (año, descripción).
6 Inserta datos y oprime aceptar.	6.1 Guarda datos.
	7 Muestra notificación y termina el caso de uso.
Flujos Alternos	
6. a Selecciona la opción Cancelar	
6.a.1 Cancela la opción de guardar.	6.a.2 Muestra mensaje de confirmación Muestra pantalla de histograma.
Poscondiciones	Se construye histograma de frecuencias de palabras claves.

Tabla 11. Descripción del caso de uso Contar palabras claves por año.

Caso de Uso:	Contar palabras claves por año.
Actores:	Experto
Resumen:	El caso de uso comienza cuando el actor selecciona la funcionalidad de frecuencia palabras claves por año.
Precondiciones:	Estar normalizadas las palabras claves.
Referencias:	RF-5.
Prioridad:	Alta.
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
1 Selecciona la opción frecuencia de palabras claves por año.	1.1 Extrae años de los documentos Muestra pantalla para seleccionar año.
2 Selecciona año.	2.1 Filtra los documentos por año. Selecciona raíz de palabras clave por documentos.
	3 Crea matriz de mxn donde m son las palabras claves y n los documentos.
	3.1 Busca palabras claves en los campos usando la raíz.
	4 Llena matriz de palabras claves x documentos.
	4.1 Crea tablas de frecuencias palabras claves, numero de aparición.
	5 Modela histogramas de las tablas de frecuencias.
5.1 Selecciona la opción de guardar.	6 Muestra pantalla (año, descripción).
6.1 Insertar los datos y acepta.	7 Guardar datos.
	7.1 Muestra notificación termina el caso de uso.
Flujos Alternos	
6. a Selecciona la opción Cancelar	
6.a.1 Cancela la opción de guardar.	6.a.2 Muestra pantalla de histograma.
Poscondiciones	Queda construido el histograma de la tabla de frecuencia de palabras claves por año.

Tabla 12. Descripción del caso de uso Contar autores por palabras claves y año.

Caso de Uso:	Contar autores por palabras claves y año.
Actores:	Experto
Resumen:	El caso de uso comienza cuando el actor selecciona la funcionalidad frecuencia de autores por palabras claves y año.
Precondiciones:	Deben estar llenos las tablas autores y documentos de patentes y artículos. Deben estar normalizados los

	autores.
Referencias:	RF-6.
Prioridad:	Alta.
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
1 Selecciona la opción frecuencia de autores por palabras claves y año.	1.1 Extrae año de los documentos. Extrae las palabras claves. Muestra pantalla con selección de año y palabras claves.
2 Selecciona año y palabra clave.	2.1 Extrae los documentos por año seleccionado.
	3 Filtra documentos por palabra clave seleccionada.
	3.1 Extrae autores de la tabla.
	4 Crea matriz (autores, documentos).
	4.1 Busca autores por documentos.
	5 Crea tablas de frecuencias absolutas.
	5.1 Modela histogramas de las tablas de frecuencias.
6 Selecciona la opción de guardar.	6.1 Muestra pantalla (año, descripción).
6 Inserta datos y oprime aceptar.	7.1 Guarda datos.
	8 Muestra notificación y termina el caso de uso.
Flujos Alternos	
6. a Selecciona la opción Cancelar	
6.a.1 Cancela la opción de guardar.	6.a.2 Muestra pantalla de histograma.
Poscondiciones	Queda construido el histograma de la tabla de frecuencia de autores por palabras claves y año.

Tabla 13. Descripción del caso de uso Modelar relaciones sociales.

Caso de Uso:	Modelar relaciones sociales.
Actores:	Experto.
Resumen:	El caso de uso comienza cuando el experto selecciona la opción de modelar los documentos.
Precondiciones:	Deben estar disponibles los metadatos de las fuentes de patentes y artículos y de la tabla autor. Deben estar normalizados los autores.
Referencias:	RF-8.
Prioridad:	Alta.
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
1. Selecciona modelar relaciones sociales.	1.1. Extrae autores de la tabla.
	2 Crea matriz binaria (documentos, autores).
	2.1 Busca autores citados por documentos.
	3 Llena matriz binaria (documentos, autores).
	3.1 Crea matriz de ocurrencia y coocurrencias (autores, autores) de la matriz anterior.
	4 Construye matriz de similitud calculando coeficiente de correlación de Pearson.
	4.1 Realiza análisis de clúster aglomerados a la matriz de similitud.
	5 Modela sociograma de autores.
6 Selecciona la opción de guardar.	6.1 Muestra pantalla (año, descripción).

7 Inserta los datos y acepta	7.1 Guarda los datos.
8 Selecciona medidas de centralidad.	8.1 Realiza cálculos de centralidad. Muestra tabla con autores con los siguientes campos. Autor, grado de salida, gado de entrada. Grado salida: Numero de autores que cita. Grado entrada: Numero de autores que lo citan.
9 Selecciona medidas de intermediación.	9.1 Realiza cálculos de intermediación. Muestra tabla de autores que permiten la conexión entre los distintos personajes de la red seria: Tabla: Autor, grado de intermediación, grado de intermediación normalizado. Grado de intermediación: Numero de pares de nodos que un actor es capaz de conectar. Grado de intermediación normalizado: Grado de intermediación expresado en porcentaje.
Flujos Alternos	
6. a Selecciona la opción Cancelar	
6.a.1 Cancela la opción de guardar.	6.a.2 Muestra pantalla de sociograma.
Poscondiciones	Quedan modeladas las relaciones sociales a través de sociogramas de los documentos.

Anexo 4. Diagramas de Clases del Análisis.

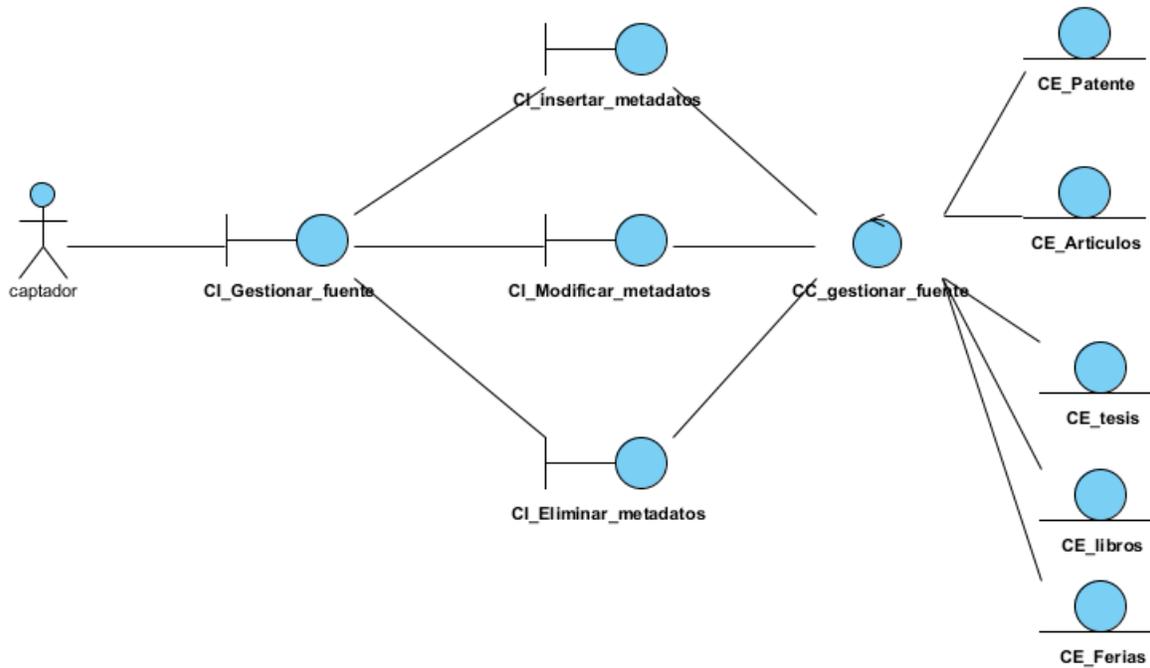


Figura 22. Diagrama de clase del análisis Gestionar fuentes.

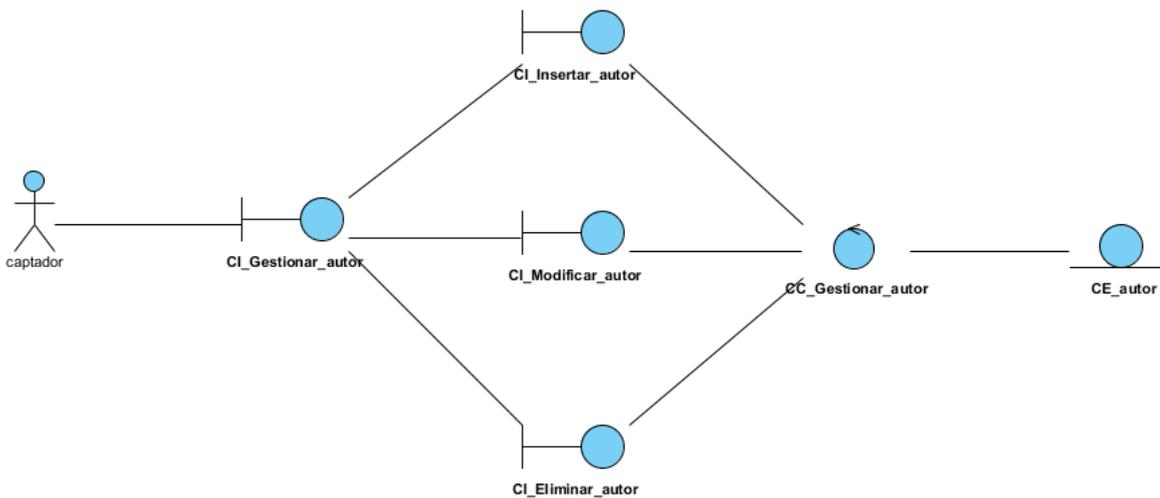


Figura 23. Diagrama de clase del análisis Gestionar autor.

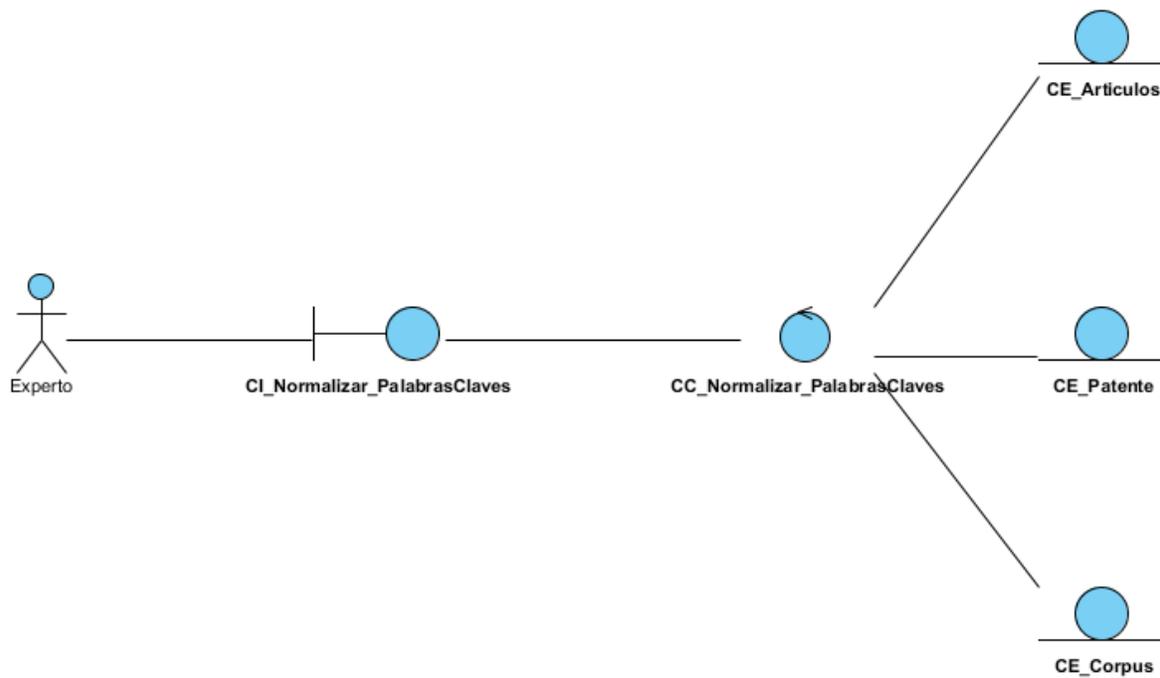


Figura 24. Diagrama de clase del análisis Normalizar palabras claves.

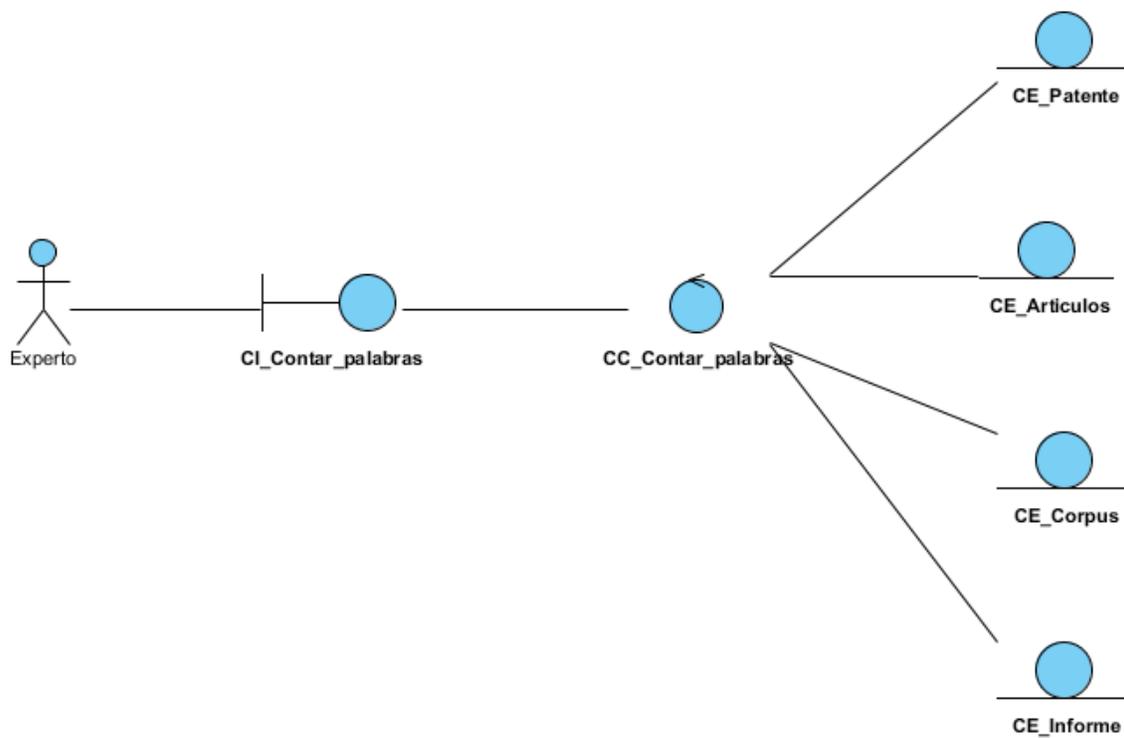


Figura 25. Diagrama de clase del análisis Contar palabras.

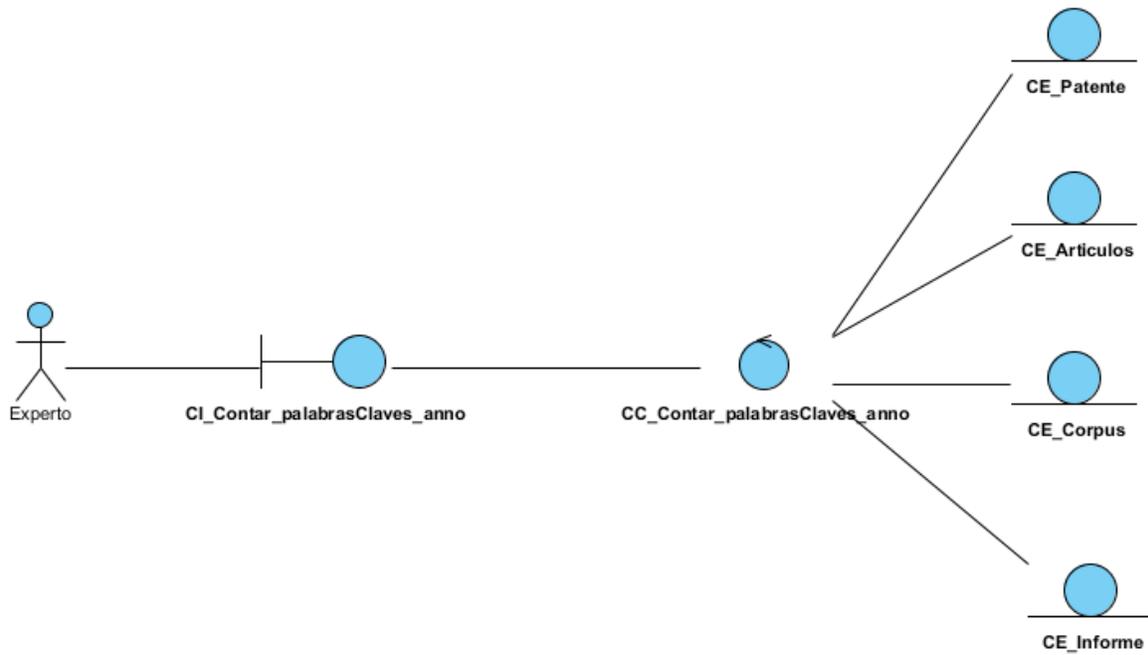


Figura 26. Diagrama de clase del análisis Contar palabras claves por año.

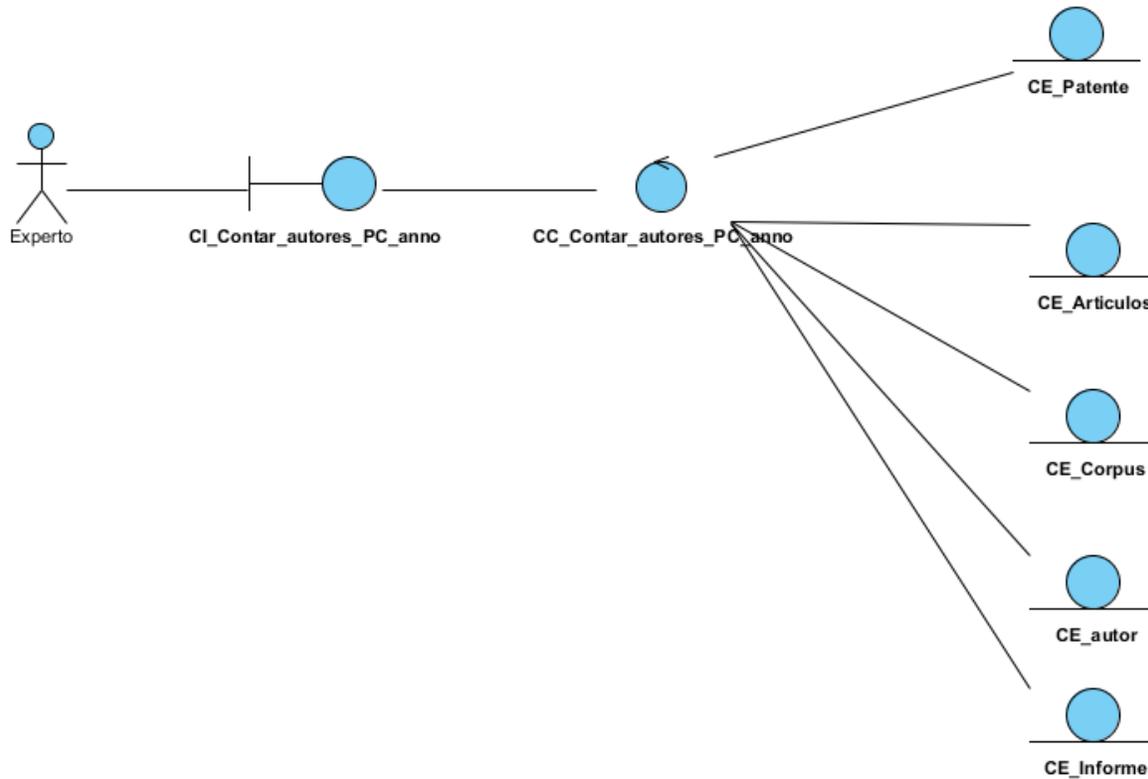


Figura 27. Diagrama de clase del análisis Contar autores por año.

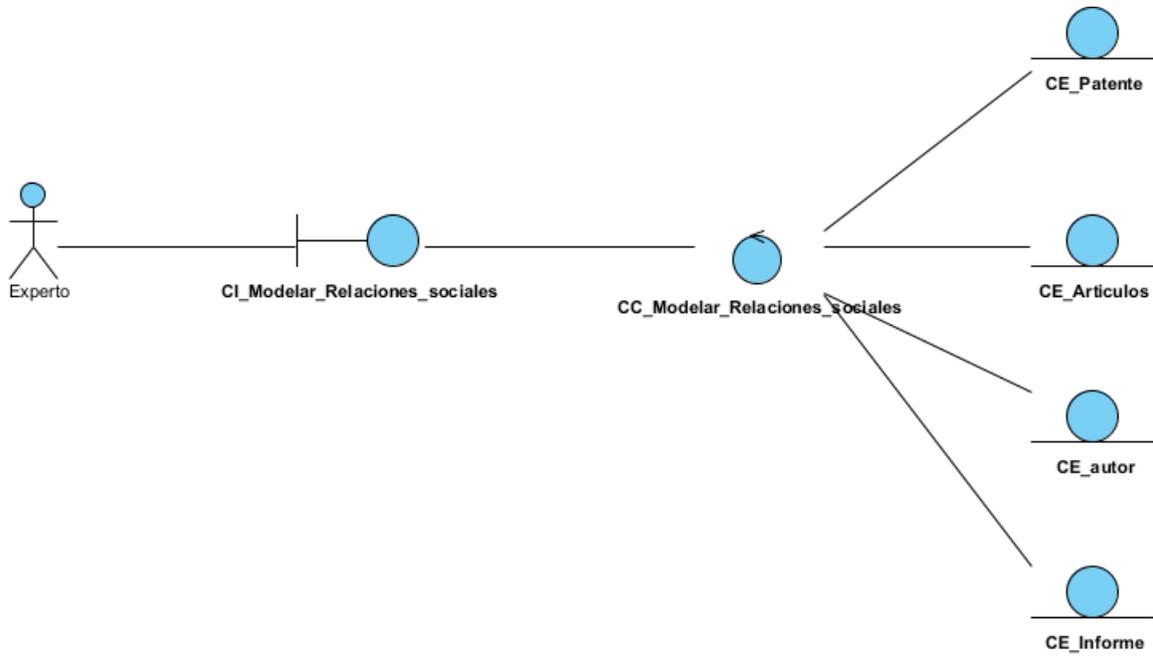


Figura 28. Diagrama de clase del análisis Modelar relaciones sociales.

Anexo 5. Diagramas de Colaboración.

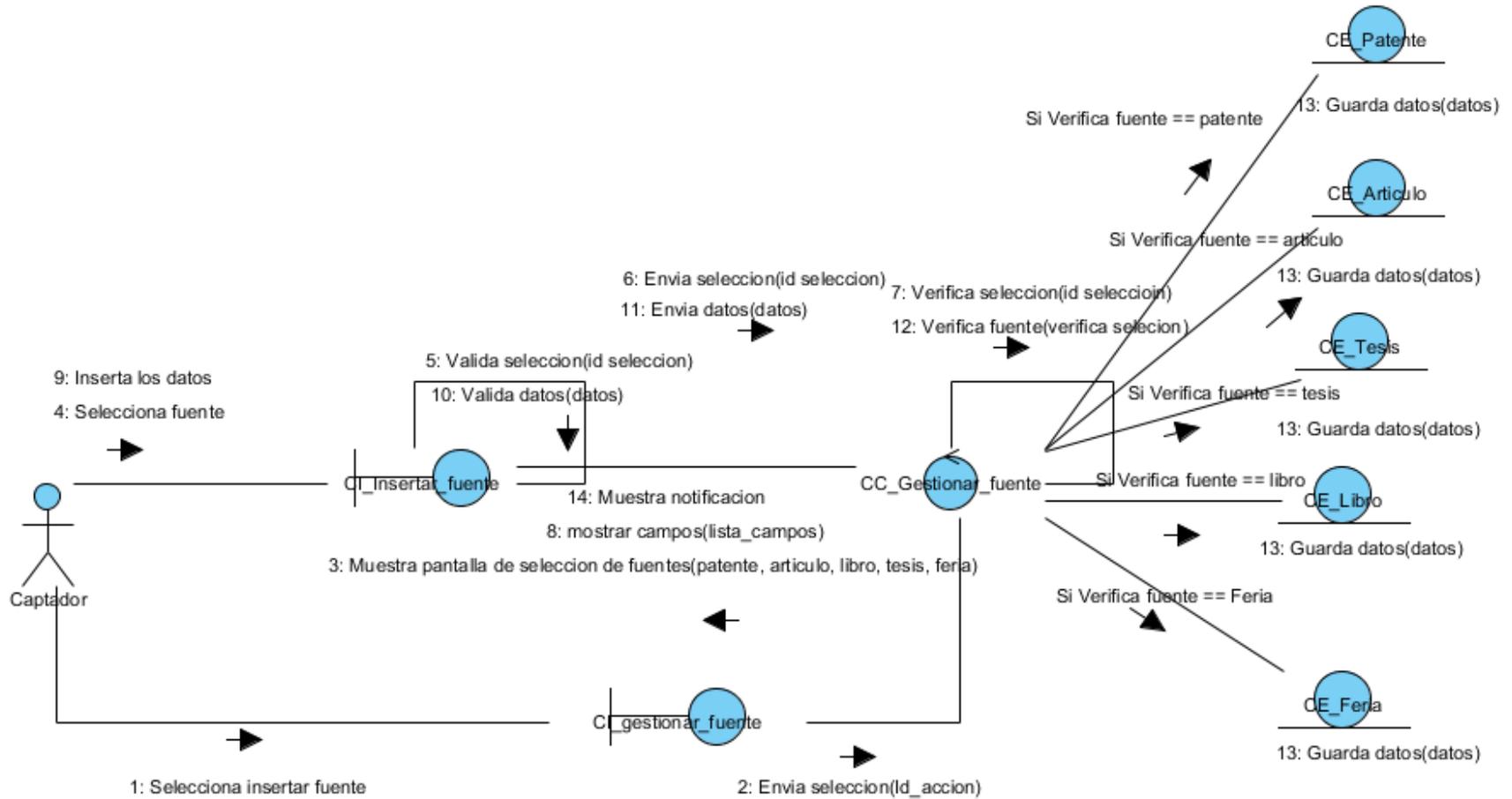


Figura 29. Diagrama de colaboración Insertar Fuentes.

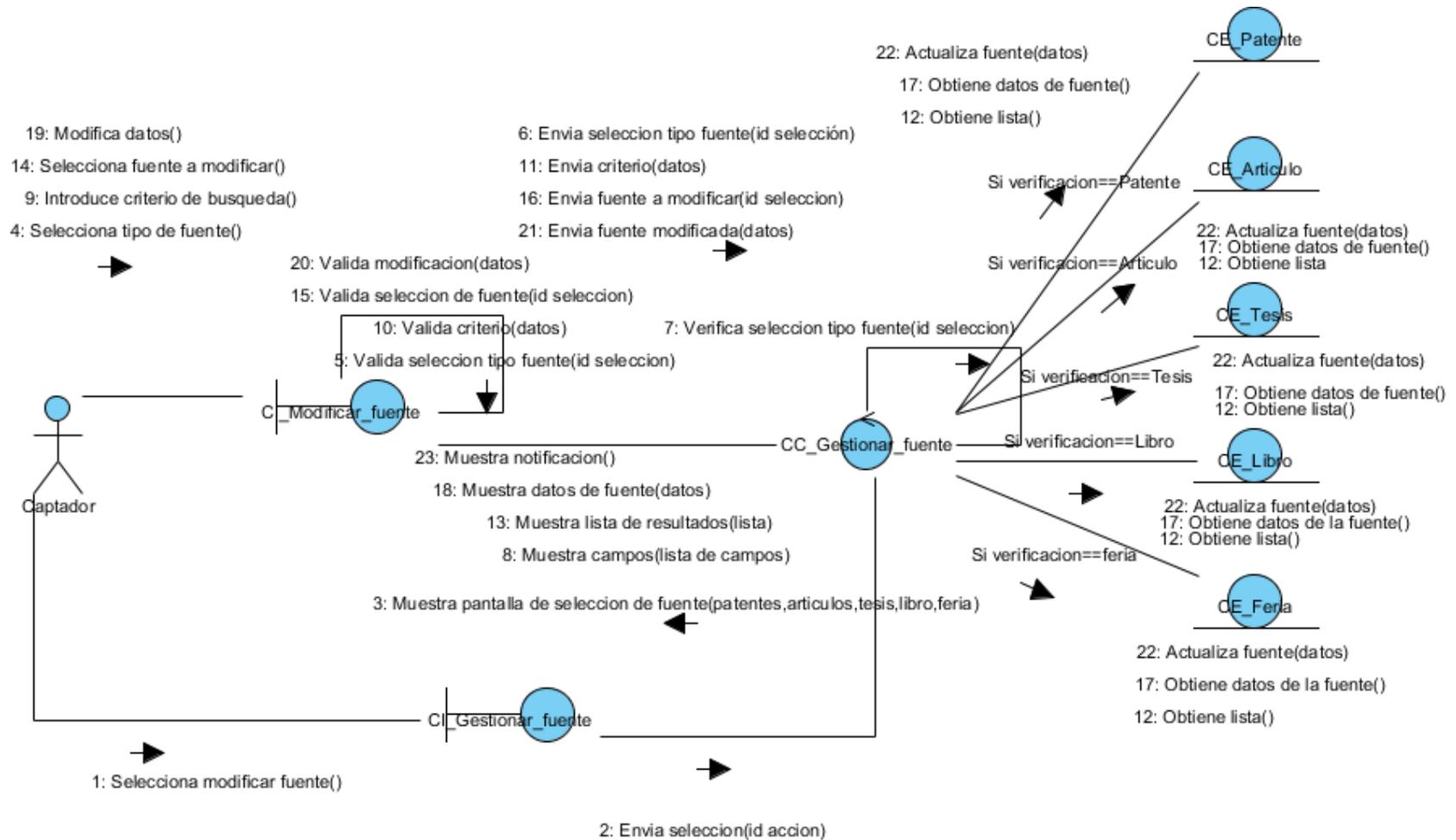


Figura 30. Diagrama de colaboración Modificar Fuentes.

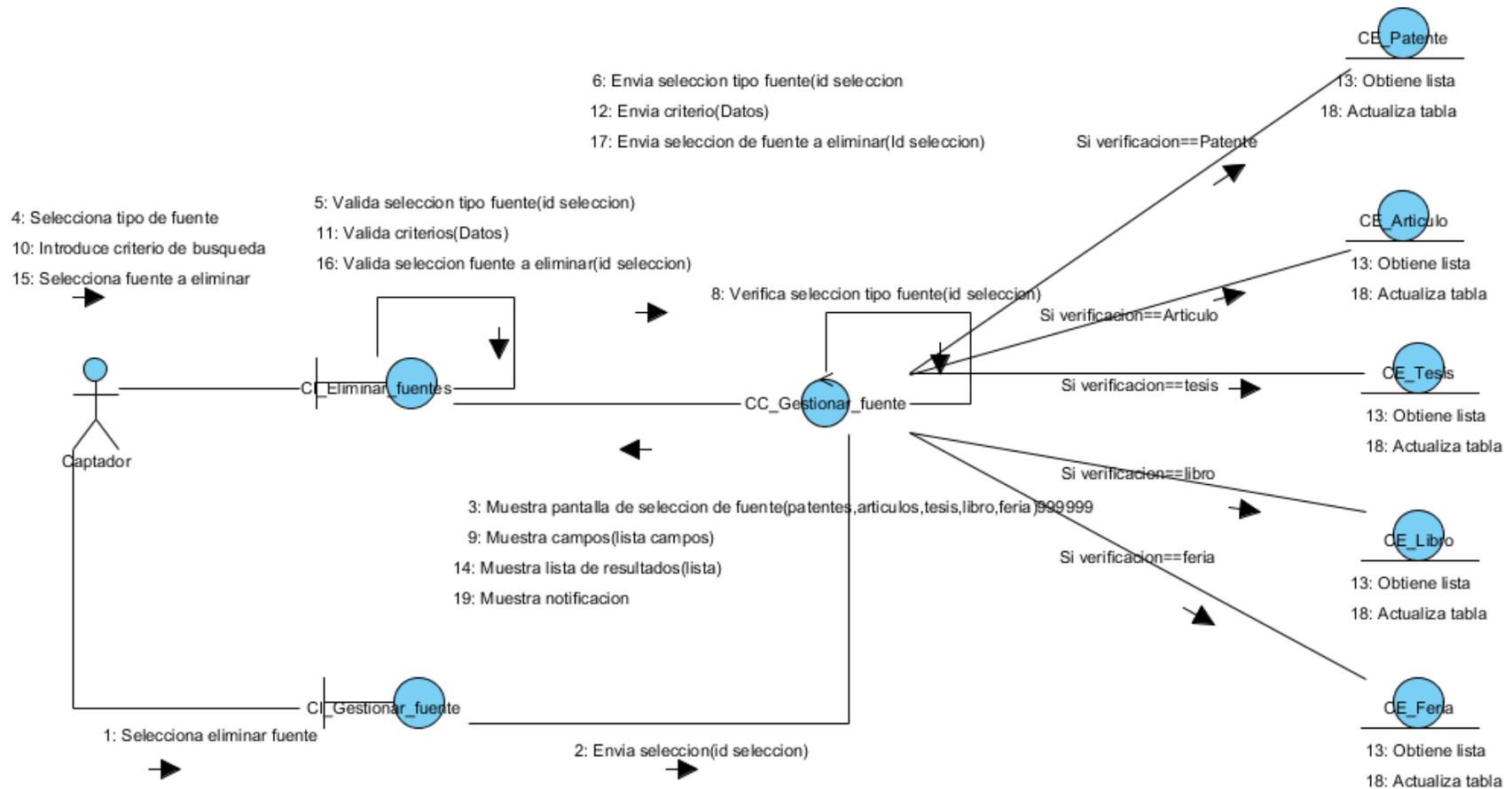


Figura 31. Diagrama de colaboración Eliminar fuentes.

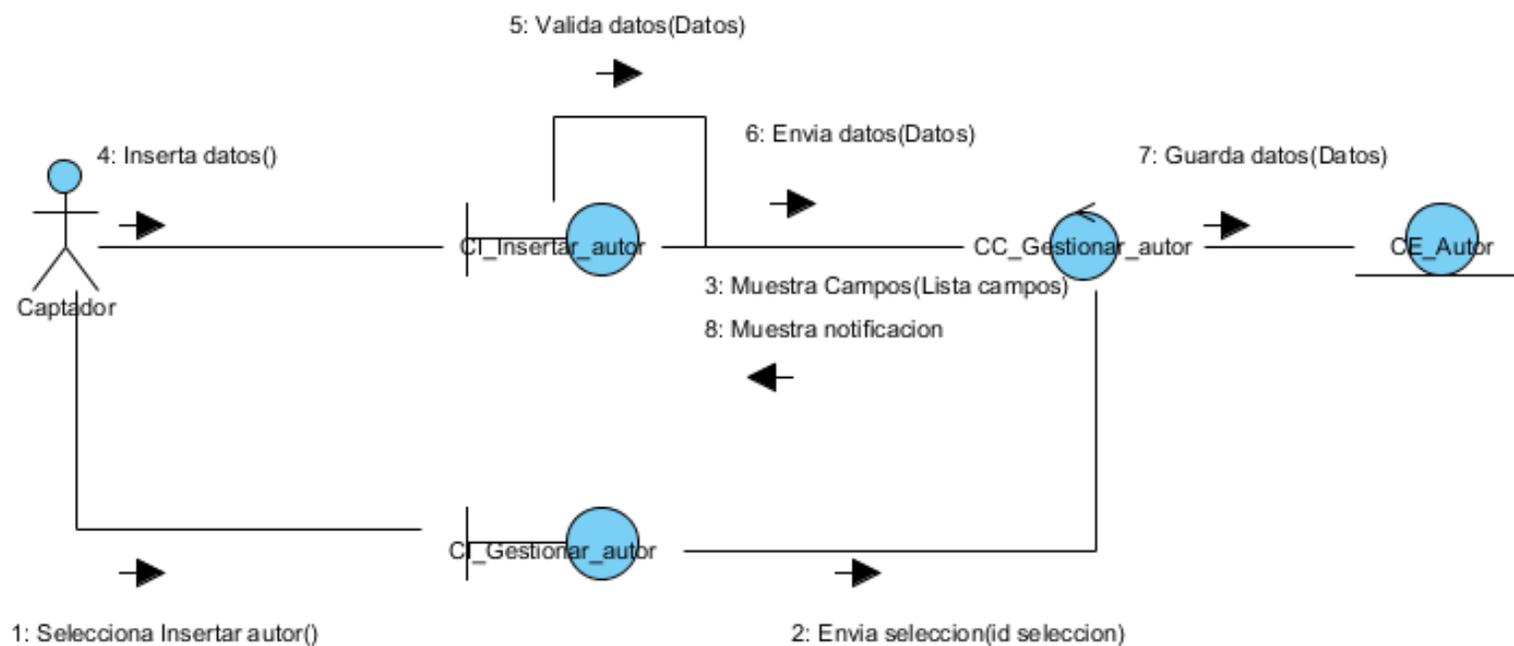


Figura 32. Diagrama de colaboración Insertar autor.

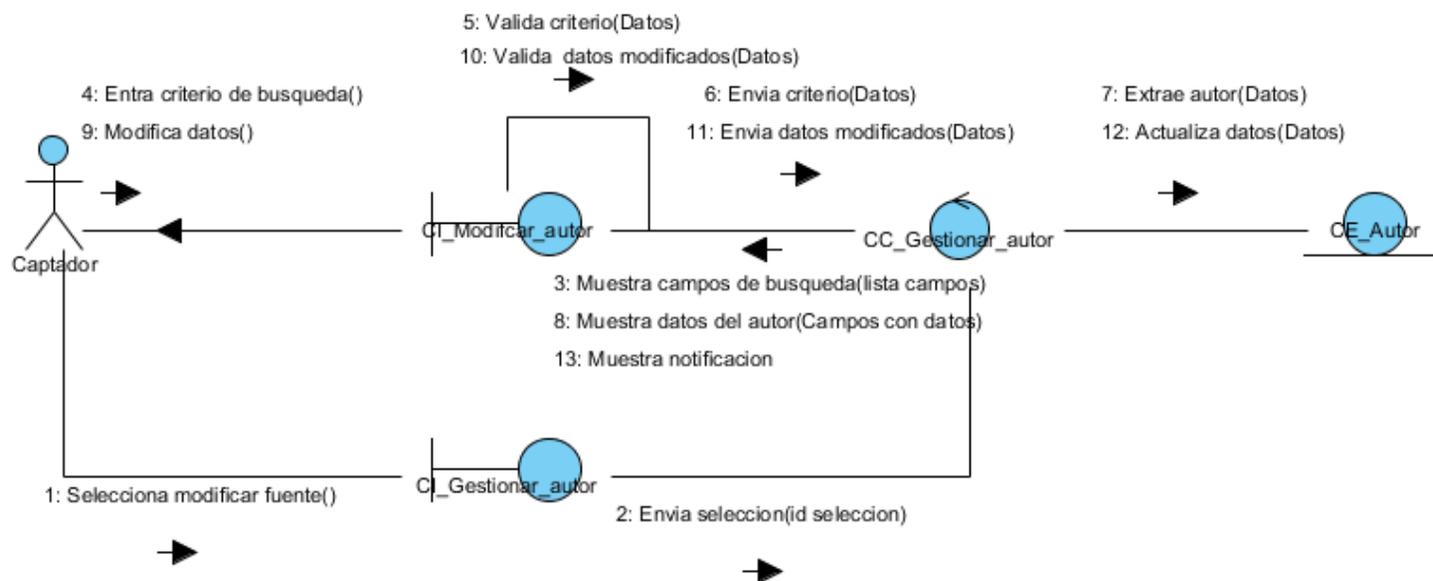


Figura 33. Diagrama de colaboración Modificar autor.

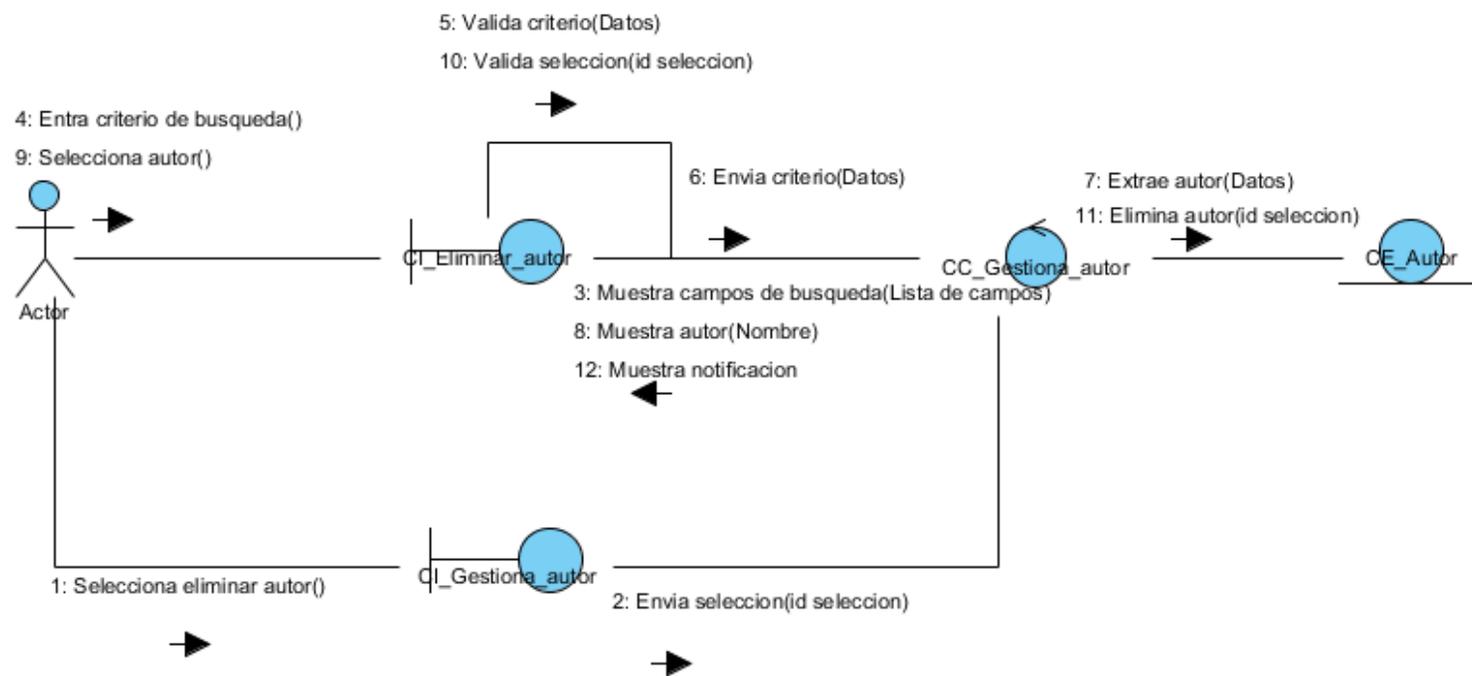


Figura 34. Diagrama de colaboración Eliminar autor.

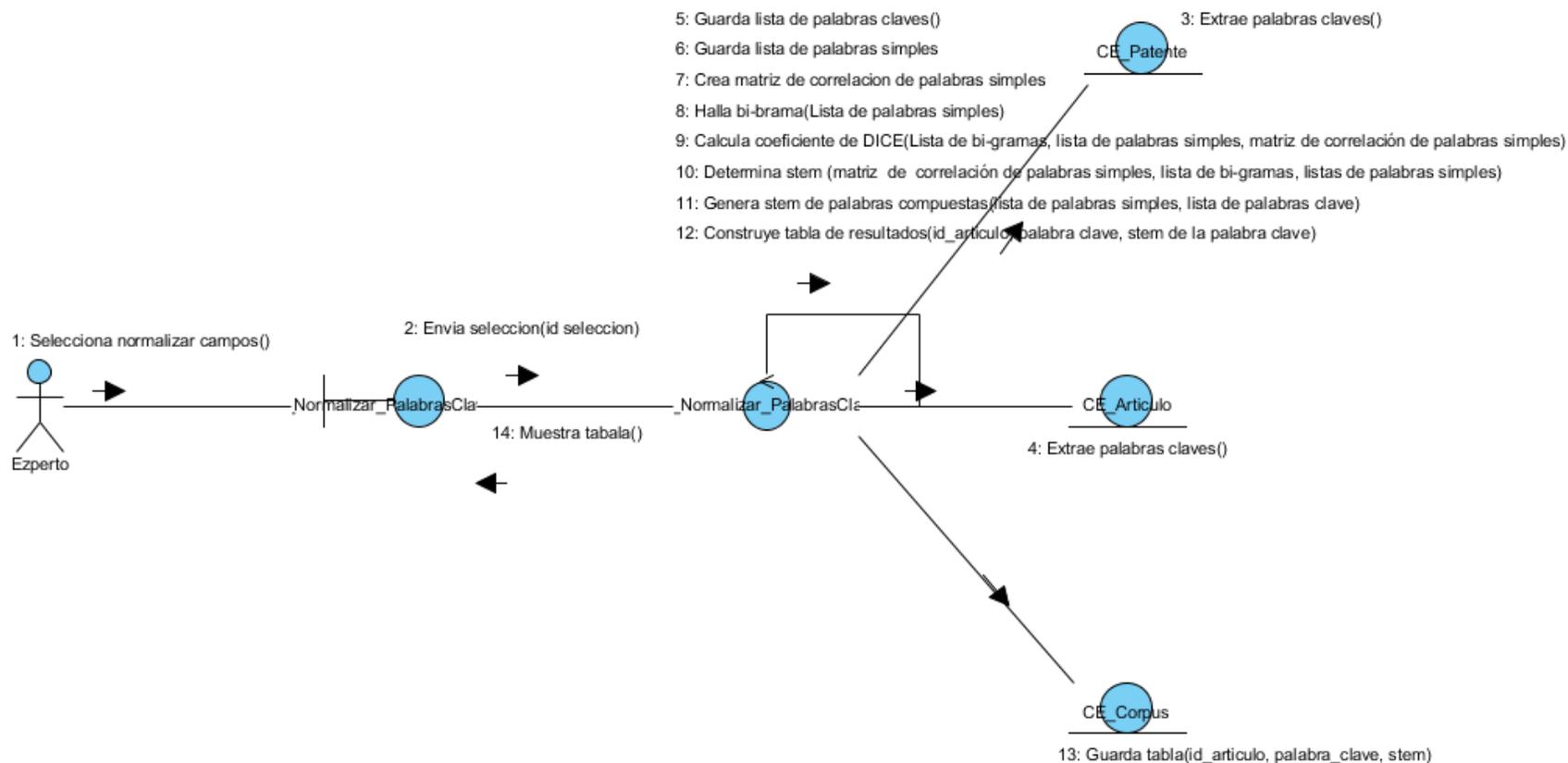


Figura 35. Diagrama de colaboración Normalizar palabras claves.

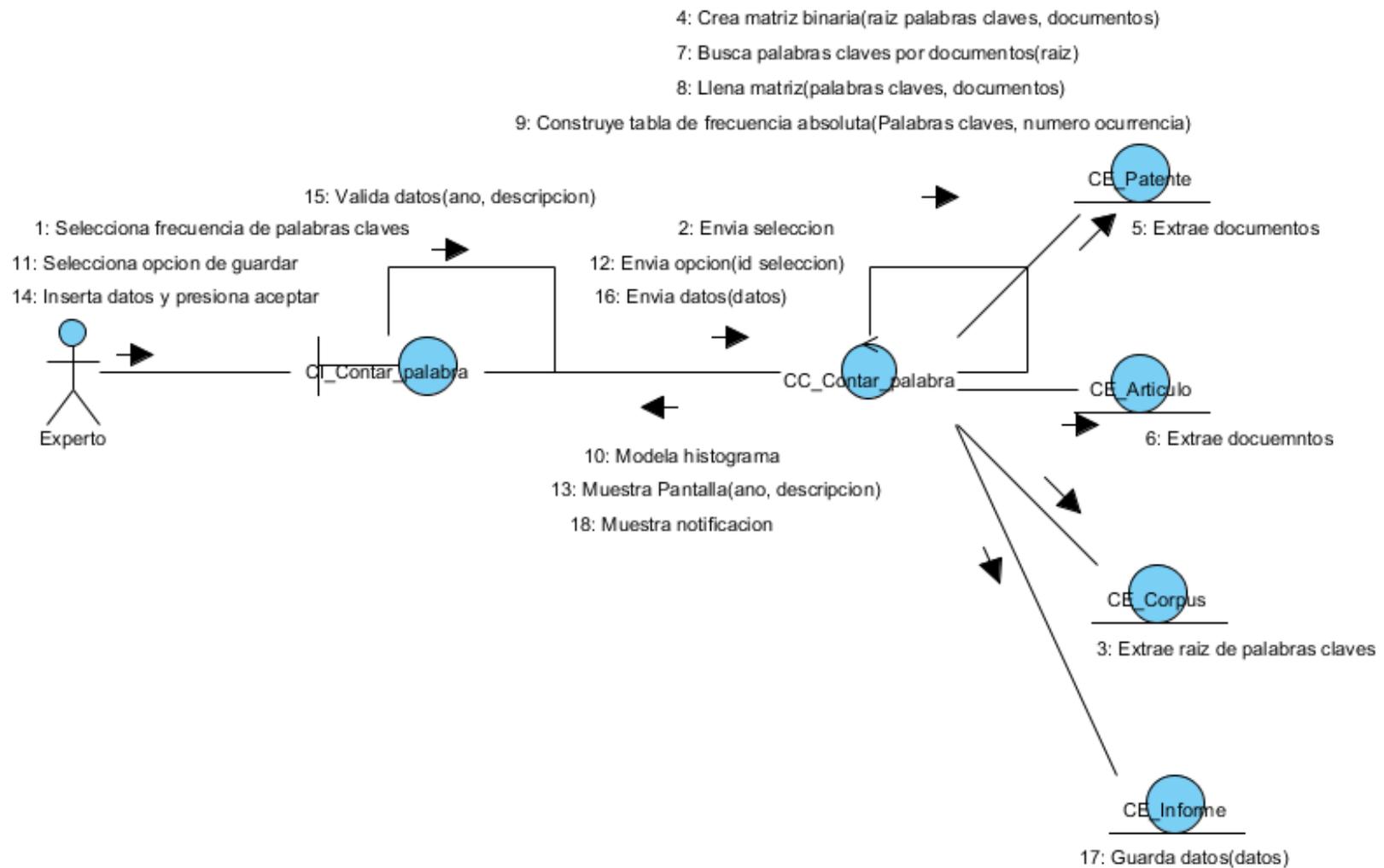


Figura 36. Diagrama de colaboración Contar palabra.

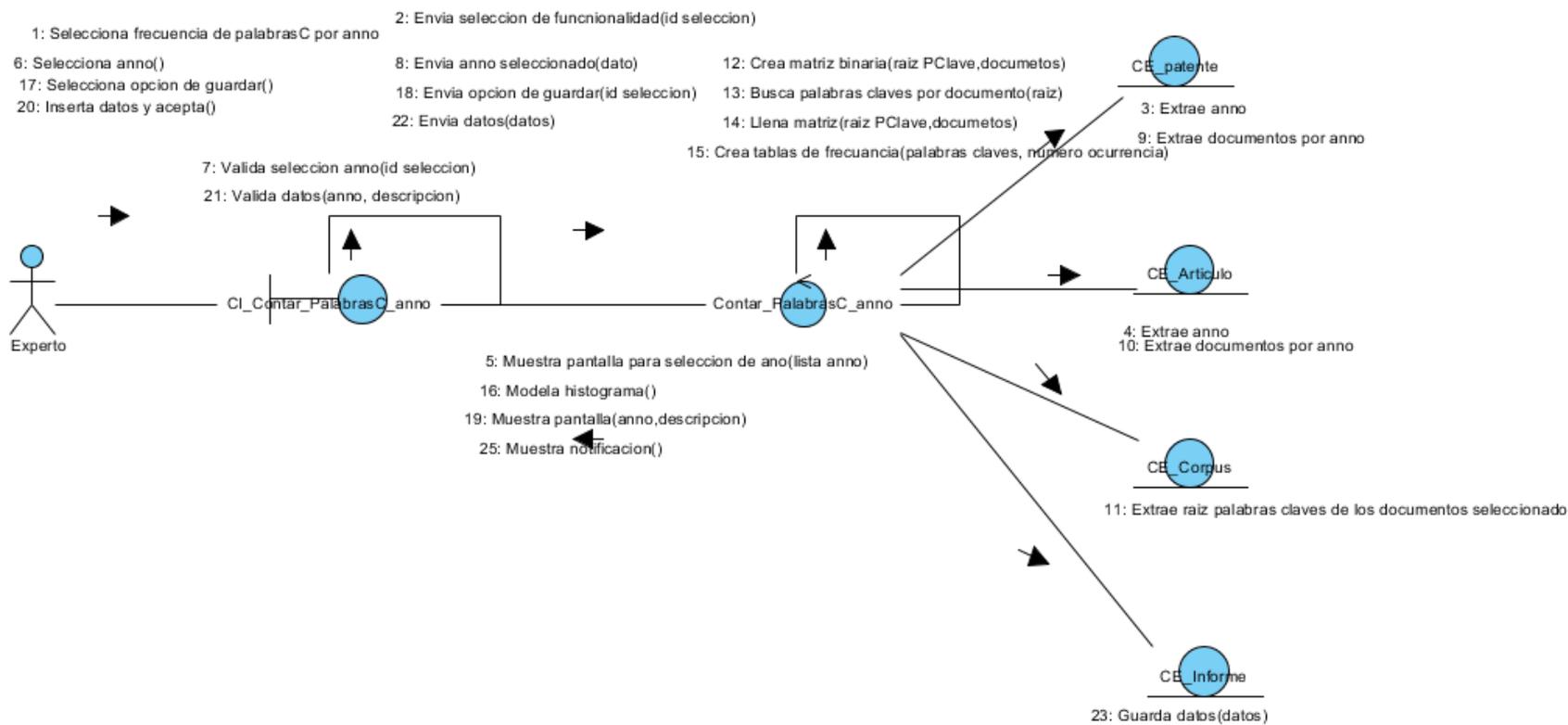


Figura 37. Diagrama de colaboración Contar palabras claves por año.

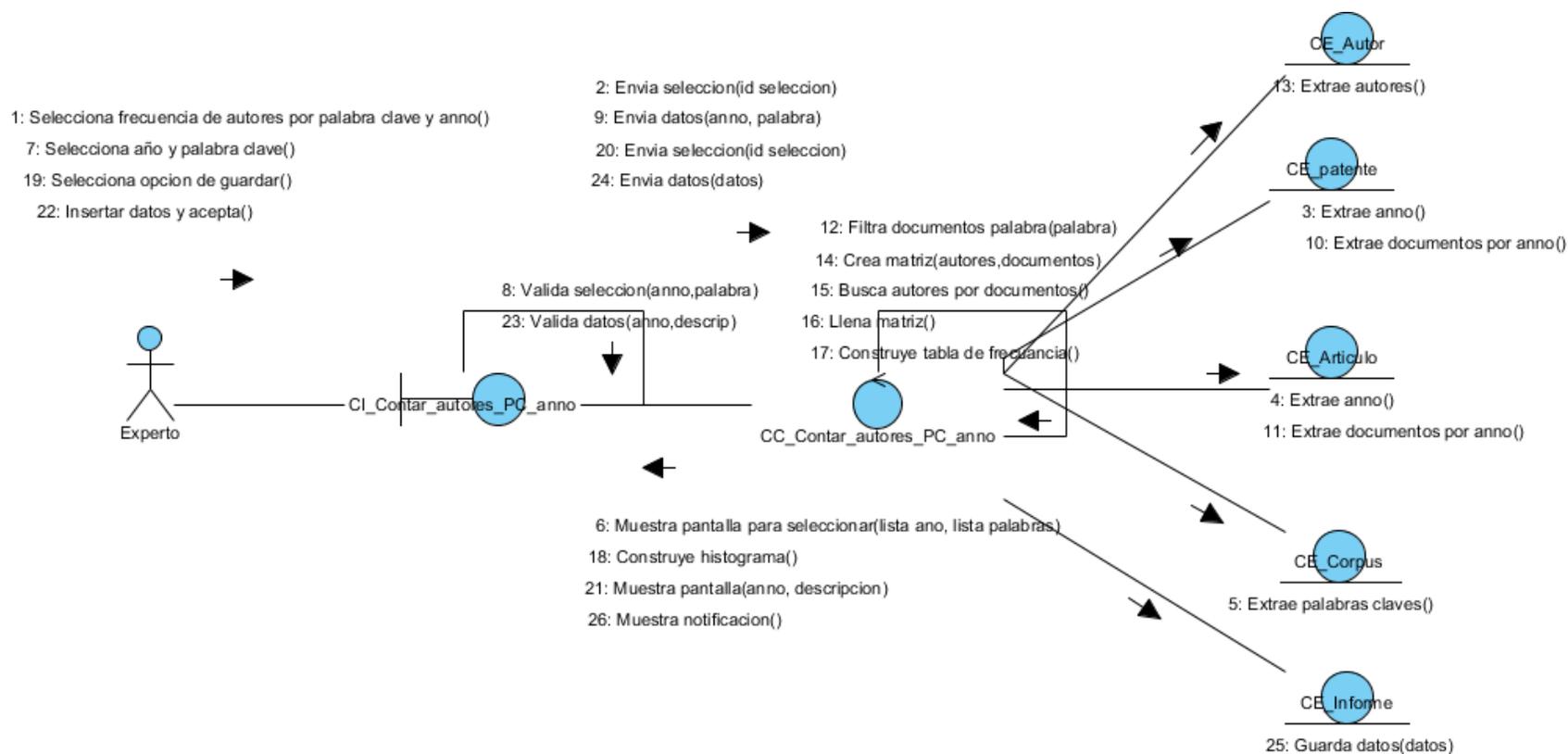


Figura 38. Diagrama de colaboración contar autor palabra clave y año.

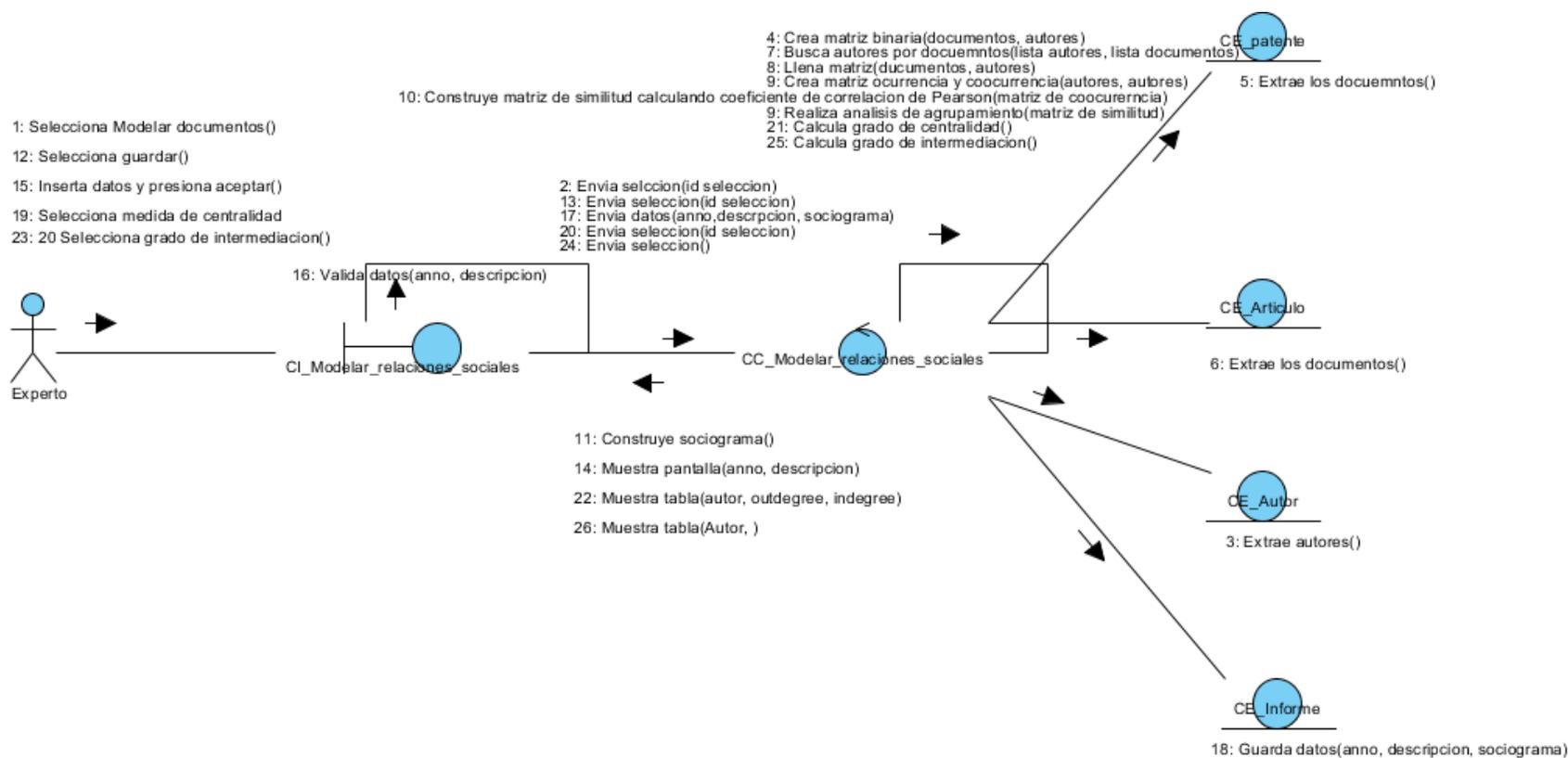


Figura 39. Diagrama de colaboración Modelar relaciones sociales.

Anexo 6. Diagramas de Clases del Diseño.

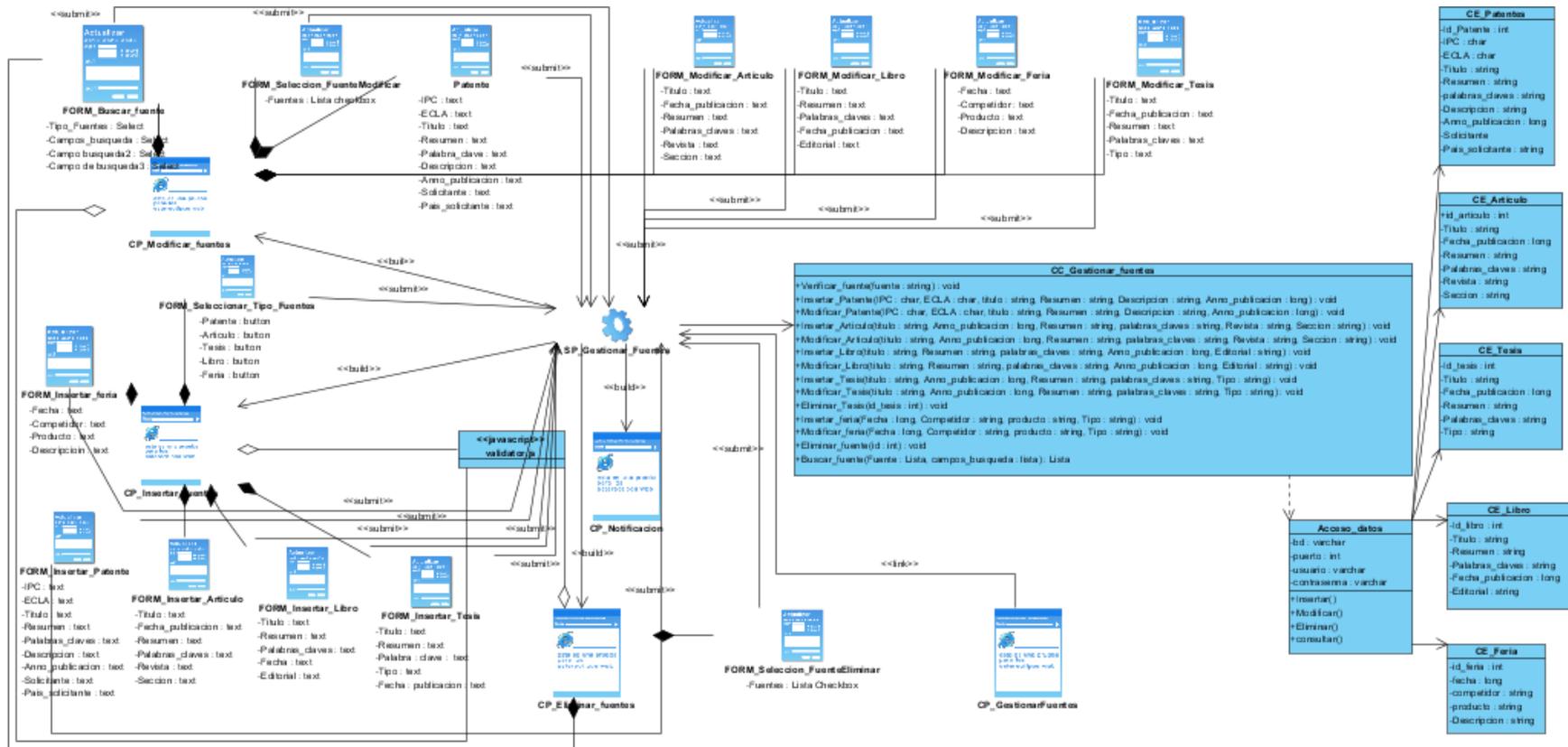


Figura 40. Diagrama de clase diseño Gestionar fuentes.

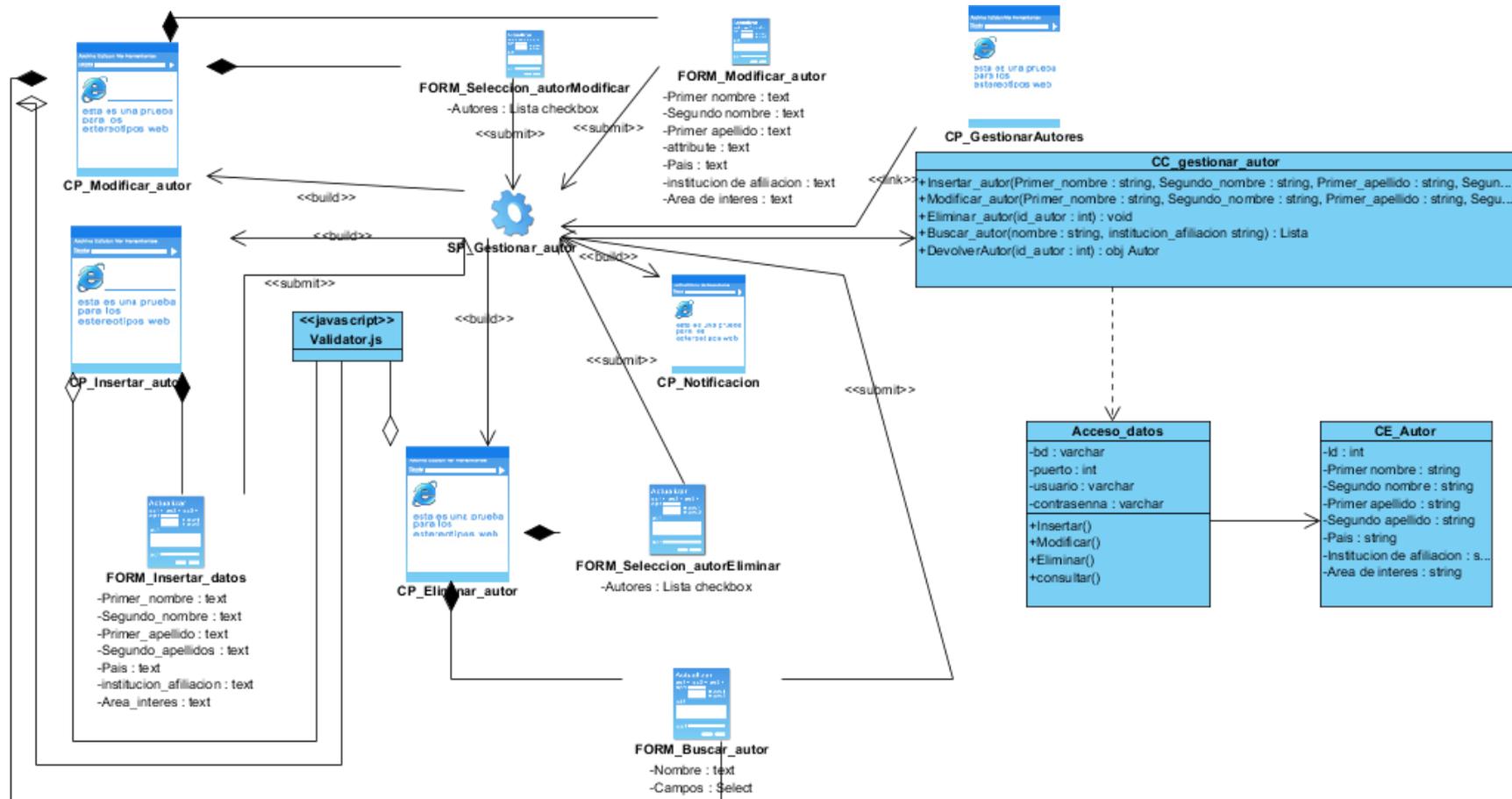


Figura 41. Diagrama de clase diseño Gestionar autor.

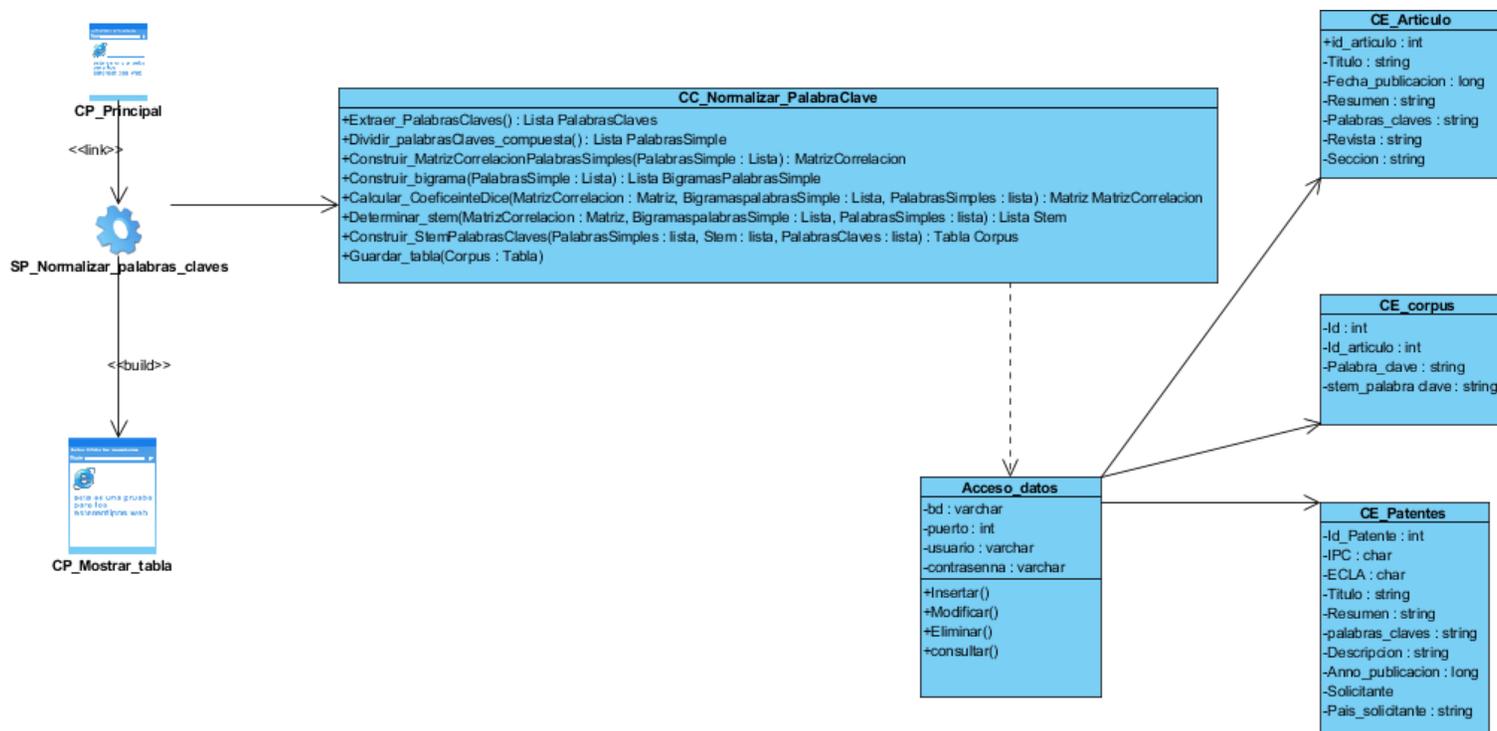


Figura 42. Diagrama de clase diseño Normalizar palabras claves.

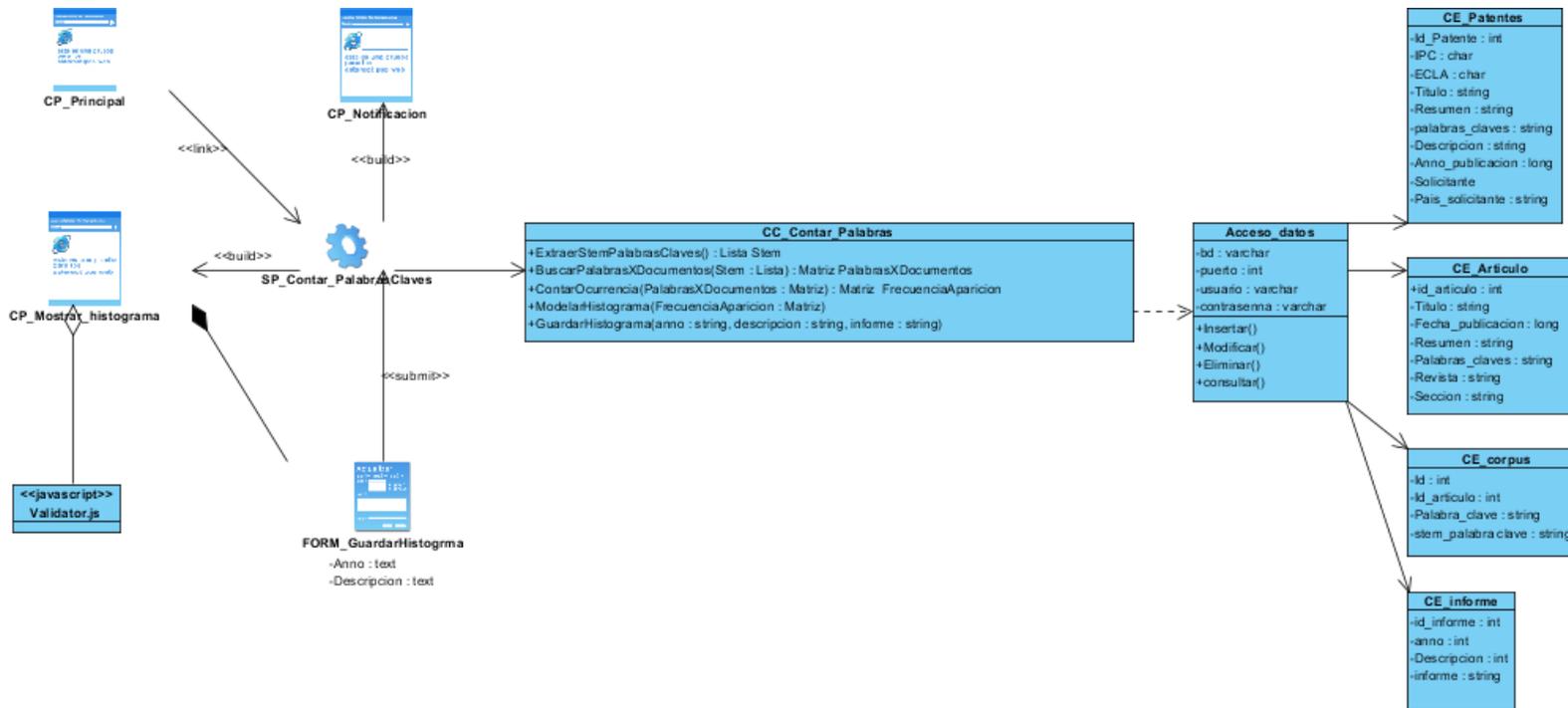


Figura 43. Diagrama de clase diseño Contar palabra.

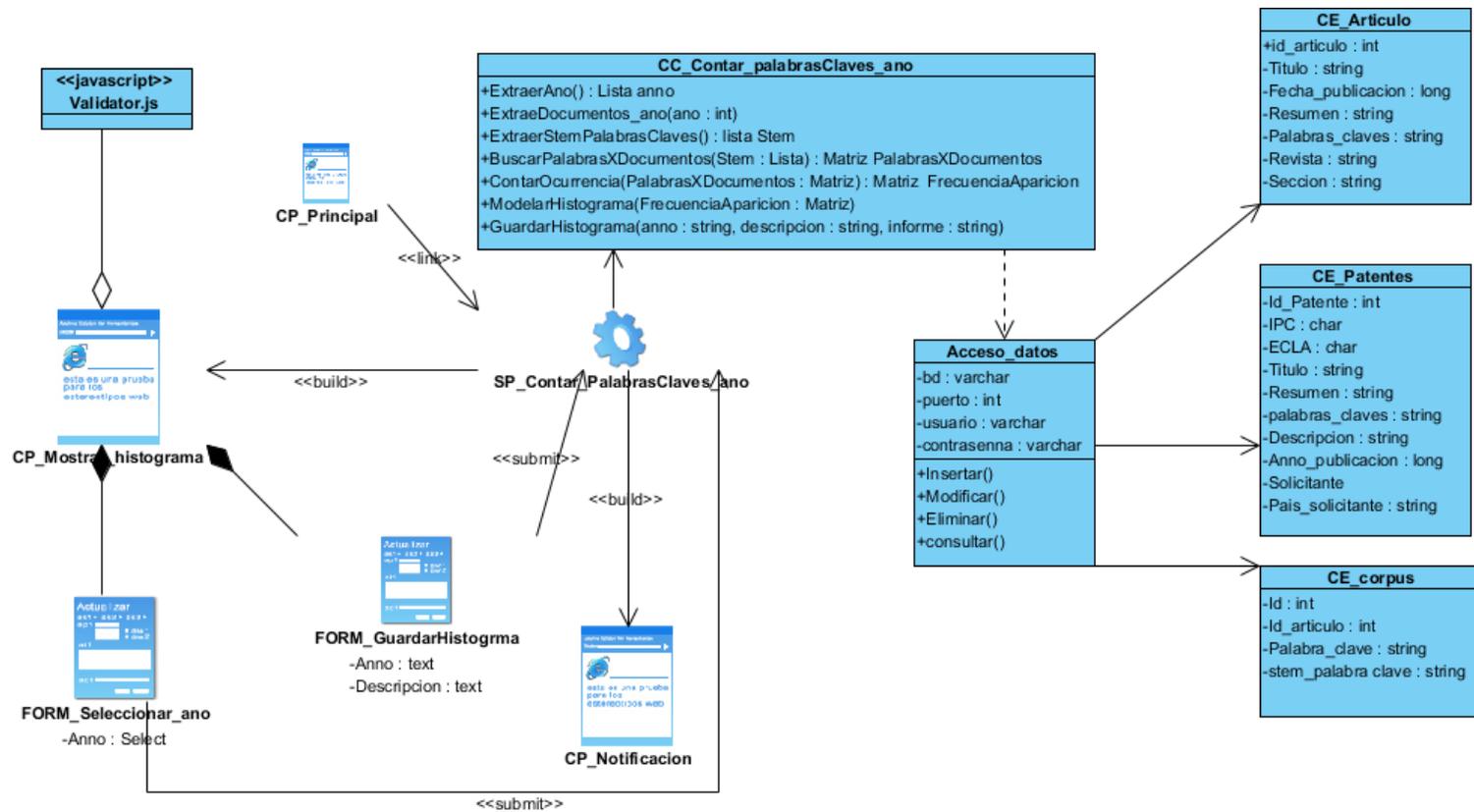


Figura 44. Diagrama de clase diseño Contar palabras claves por año.

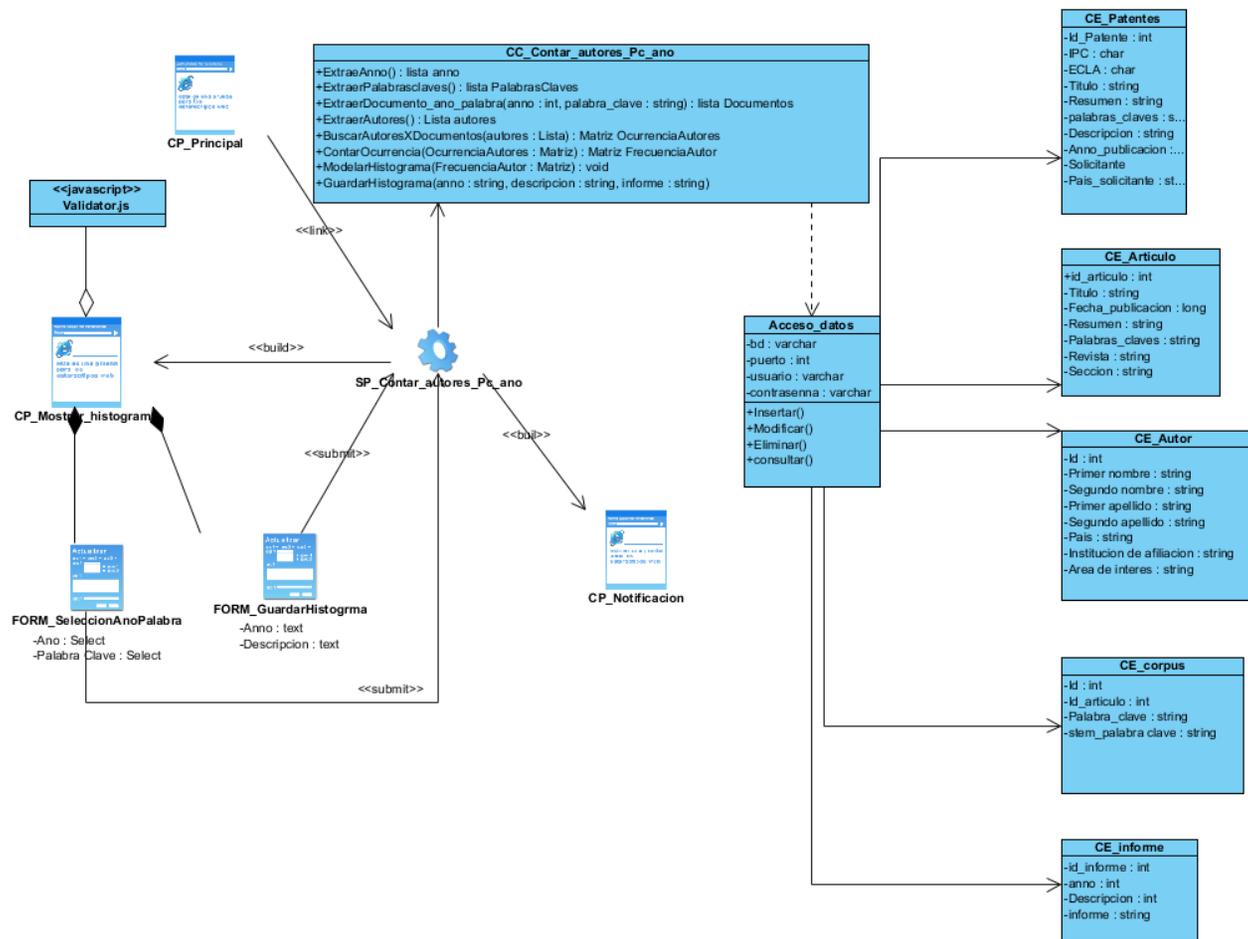


Figura 45. Diagrama de clase diseño Contar autores por palabra clave y año.

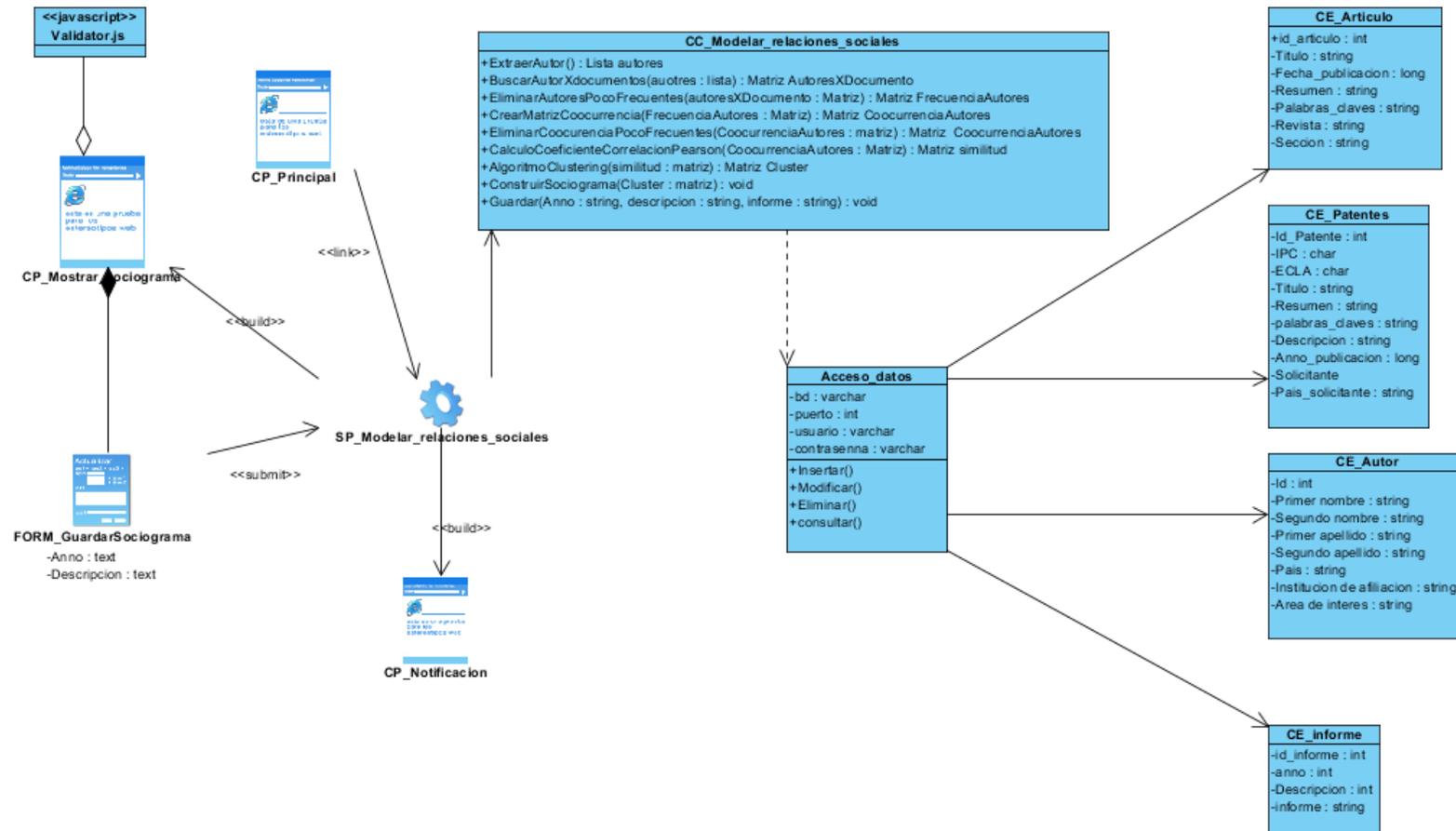


Figura 46. Diagrama de clase diseño Modelar relaciones sociales.

Anexo 7. Resultados de la validación de los requerimientos.

Revisores

R1. Ing. Celia María Soulyar Reyes (Experta con 5 años de experiencia).

R2. Ms. Ismael Armando Nodarse Mora (Ingeniería de Software).

R3. Grupo Calidad Facultad 4.

Los requerimientos según los indicadores fueron clasificados por cada revisor en:

C: Correcto.

I: Incorrecto.

El resultado final (**Resultado**) será C o I; para esto deben de coincidir al menos dos revisores atendiendo al indicador.

Anexo 7.1. Corrección

Tabla 14. Resultado de la Corrección.

Requerimientos	Revisores			Resultado
	R ₁	R ₂	R ₃	
RF ₁	C	C	C	C
RF ₂	C	C	C	C
RF ₃	C	I	I	I
RF ₄	C	C	C	C
RF ₅	C	C	C	C
RF ₆	C	C	C	C
RF ₇	C	C	C	C
RF ₈	C	C	C	C
RNF ₁	C	C	C	C
RNF ₂	C	C	C	C
RNF ₃	C	C	C	C
RNF ₄	C	C	C	C
RNF ₅	C	C	C	C
RNF ₆	C	C	C	C
RNF ₇	C	C	C	C
RNF ₈	C	C	C	C
RNF ₉	C	C	C	C
RNF ₁₀	C	C	C	C
RNF ₁₁	C	C	C	C
Total correctos				18
Total incorrectos				1

Anexo 7.2. Compleción

Tabla 15. Resultado de la Compleción

Requerimientos	Revisores			Resultado
	R ₁	R ₂	R ₃	
RF ₁	C	C	C	C
RF ₂	C	C	C	C
RF ₃	C	C	C	C
RF ₄	C	C	C	C
RF ₅	C	C	C	C
RF ₆	C	C	C	C
RF ₇	C	C	C	C
RF ₈	C	C	C	C
RNF ₁	C	C	C	C
RNF ₂	C	C	C	C
RNF ₃	C	C	C	C
RNF ₄	C	C	C	C
RNF ₅	C	C	C	C
RNF ₆	C	C	C	C
RNF ₇	C	C	C	C
RNF ₈	C	C	C	C
RNF ₉	C	C	C	C
RNF ₁₀	C	C	C	C
RNF ₁₁	C	C	C	C
Total correctos				19
Total incorrectos				0

Anexo 7.3. Comprensión

Tabla 16. Resultado de la Comprensión

Requerimientos	Revisores			Resultado
	R ₁	R ₂	R ₃	
RF ₁	C	I	C	C
RF ₂	C	C	C	C
RF ₃	C	I	C	C
RF ₄	C	C	C	C
RF ₅	C	C	C	C
RF ₆	C	C	C	C
RF ₇	C	C	C	C
RF ₈	C	C	C	C
RNF ₁	C	C	C	C
RNF ₂	C	C	C	C
RNF ₃	C	C	C	C
RNF ₄	C	C	C	C
RNF ₅	C	C	C	C
RNF ₆	C	C	C	C

RNF ₇	C	C	C	C
RNF ₈	C	C	C	C
RNF ₉	C	C	C	C
RNF ₁₀	C	C	C	C
RNF ₁₁	C	C	C	C
Total correctos				19
Total incorrectos				0

Anexo 7.4. Consistencia interna

Tabla 17. Resultado de la Consistencia interna.

Requerimientos	Revisores			Resultado
	R ₁	R ₂	R ₃	
RF ₁	C	C	C	C
RF ₂	C	C	C	C
RF ₃	C	C	C	C
RF ₄	C	C	C	C
RF ₅	C	C	C	C
RF ₆	C	C	C	C
RF ₇	C	C	C	C
RF ₈	C	C	C	C
RNF ₁	C	C	C	C
RNF ₂	C	C	C	C
RNF ₃	C	C	C	C
RNF ₄	C	C	C	C
RNF ₅	C	C	C	C
RNF ₆	C	C	C	C
RNF ₇	C	C	C	C
RNF ₈	C	C	C	C
RNF ₉	C	C	C	C
RNF ₁₀	C	C	C	C
RNF ₁₁	C	C	C	C
Total correctos				19
Total incorrectos				0

Anexo 7.5. Consistencia externa

Tabla 18. Resultado de la Consistencia externa.

Requerimientos	Revisores			Resultado
	R ₁	R ₂	R ₃	
RF ₁	C	C	C	C
RF ₂	C	C	C	C
RF ₃	C	C	C	C
RF ₄	C	C	C	C
RF ₅	C	C	C	C
RF ₆	C	C	C	C
RF ₇	C	C	C	C

RF ₈	C	C	C	C
RNF ₁	C	C	C	C
RNF ₂	C	C	C	C
RNF ₃	C	C	C	C
RNF ₄	C	C	C	C
RNF ₅	C	C	C	C
RNF ₆	C	C	C	C
RNF ₇	C	C	C	C
RNF ₈	C	C	C	C
RNF ₉	C	C	C	C
RNF ₁₀	C	C	C	C
RNF ₁₁	C	C	C	C
Total correctos				19
Total incorrectos				0

Anexo 8. Cálculo del índice de Jaccard

Fórmula 1. Índice de Jaccard

$$J(ij) = \frac{c_{ij}}{c_i + c_j - c_{ij}} \quad \text{en donde}$$

C_{ij} = Número de ocurrencias de la palabra i e j,
 C_i = Frecuencia de la palabra i,
 C_j = Frecuencia de la palabra j

Este índice es la razón entre el número de documentos que tienen dos palabras clave y la suma total de los documentos en que éstas palabras se encuentran por separado menos los documentos en que se encuentren las dos palabras. Esta relación siempre da resultados entre 0 y 1, y es igual a 1 (uno) cuando dos palabras aparecen solamente en los mismos documentos ($C_i = C_j = C_{ij}$), y toma el valor de cero cuando dos palabras no aparecen citadas en forma conjunta en ningún documento ($C_{ij} = 0$).

Anexo 9. Fórmula de Coeficiente de DICE

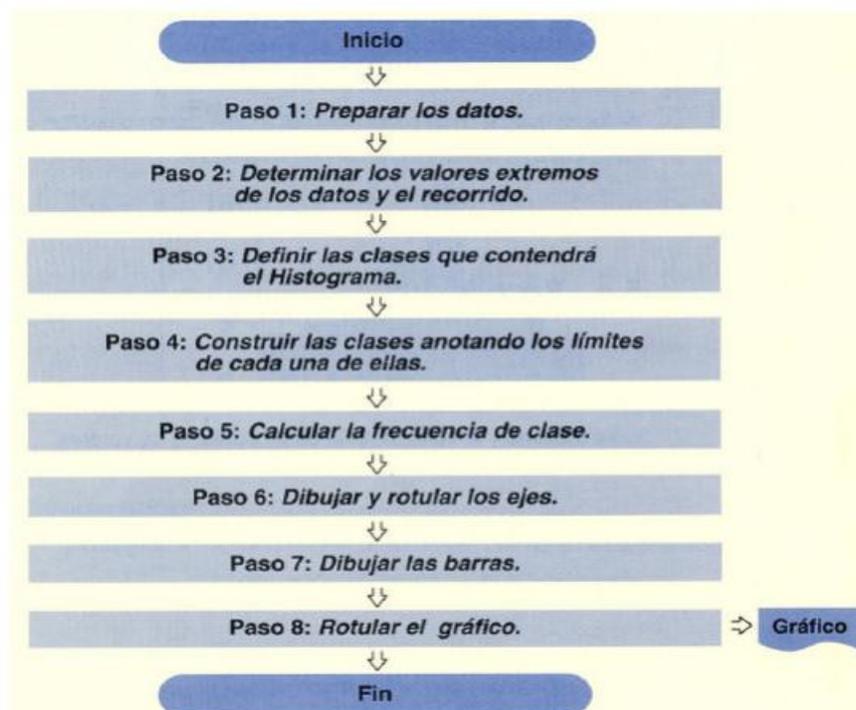
Calcular la similaridad (Dice)

$$S = \frac{2C}{A+B}$$

- A: Bi-gramas unicos de la primera palabra
- B: Bi-gramas unicos de 2a palabra
- C: Bi-gramas unicos compartidos

- Calcular S para todos los bi-gramas

Anexo 10. Histogramas



Anexo 11. Aval de calidad

**Universidad de las Ciencias Informáticas
Grupo de Calidad del Centro FORTES**

La Habana, 6 de junio de 2012
Año 54 de la Revolución

AVAL

A quien pueda interesar:

Hacemos constar que los requerimientos del Sistema de Vigilancia Tecnológica para la tecnología educativa fueron sometidos a un proceso de revisión por parte del Grupo de Calidad del Centro FORTES, teniendo en cuenta las métricas para la Calidad de la Especificación de los Requisitos basado en la consistencia de la interpretación de los revisores para cada requerimiento a través de las siguientes características: especificidad, corrección, completión, comprensión, consistencia interna y externa.

En el proceso de revisión participaron 6 revisores, que emitieron sus criterios sobre 8 requisitos funcionales y 11 no funcionales del sistema que se propone.

El resultado de la revisión fue satisfactorio por tanto se considera que los requisitos están descritos de forma clara, concisa y que cumplen con las características de las métricas para la Calidad de la Especificación de los Requisitos.



Ing. Yusdel Meriño Almaguer
Líder del Grupo de Calidad del Centro FORTES