

**MODELO SEMÁNTICO DE EXPANSIÓN DE CONSULTA PARA LA
BÚSQUEDA WEB**

ANEXOS



PROYECTO DE GRADO

**IVÁN DARÍO LÓPEZ GÓMEZ
CARLOS ADRIÁN ANDRADE HOYOS**

Director: PhD (student). Miguel Ángel Niño Zambrano

Asesor: PhD (c). Carlos Cobos Lozada

**UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES
DEPARTAMENTO DE SISTEMAS
POPAYÁN**

2011

TABLA DE CONTENIDO

1	ANEXO A.....	1
1.1	FICHAS BIBLIOGRÁFICAS RECUPERACIÓN DE INFORMACIÓN.....	1
1.2	FICHAS BIBLIOGRÁFICAS WEB SEMÁNTICA.....	6
1.3	FICHAS BIBLIOGRÁFICAS RECUPERACIÓN DE INFORMACIÓN Y WEB SEMÁNTICA	11
2	ANEXO B.....	18
2.1	ARTÍCULO DE INVESTIGACIÓN	18
3	ANEXO C.....	28
3.1	ELECCIÓN DE LA ONTOLOGÍA.....	28
3.1.1	<i>Licencia Ontología NCI.....</i>	<i>29</i>
4	ANEXO D	32
4.1	ARQUITECTURA DE LA APLICACIÓN	32
4.1.1	<i>Vistas de la Arquitectura</i>	<i>32</i>
4.1.1.1	Vista Conceptual	32
4.1.1.1.1	Diagrama de Casos de Uso	32
4.1.1.1.2	Casos de Uso en formato extendido	33
4.1.1.2	Vista Lógica.....	35
4.1.1.3	Vista Física.....	35
4.1.1.4	Vista de Implementación	36
4.1.2	<i>Patrones utilizados</i>	<i>37</i>
5	ANEXO E.....	39
5.1	PRUEBA DE RENDIMIENTO	39
5.1.1	<i>Software y Hardware utilizados para las pruebas</i>	<i>39</i>
5.2	PRUEBA DE USABILIDAD	41
6	ANEXO F.....	44
6.1	VALIDACIÓN DEL PROTOTIPO.....	44
6.1.1	<i>Curva Precisión-Recall.....</i>	<i>44</i>
6.1.2	<i>Precisión en los K primeros resultados.....</i>	<i>59</i>
6.1.2.1	Precision At k - MSEC Web Search	60
6.1.2.2	Precision At k - Google	61
6.1.2.3	Precision At k - Yahoo!	63

6.1.3	<i>Estadísticas Kappa</i>	65
6.1.4	<i>Índice MAP de GoPubMed y MSEC Web Search</i>	68
7	ANEXO G	70
7.1	MANUAL DE USUARIO	70
7.1.1	<i>Ingresar al sistema</i>	70
7.1.2	<i>Registrar Usuario</i>	70
7.1.3	<i>Búsqueda Web</i>	71
7.1.4	<i>Obtener Ayuda</i>	72
7.1.5	<i>Cerrar Sesión</i>	73
8	REFERENCIAS	74

ÍNDICE DE TABLAS

TABLA 1. FICHA B 1 RECUPERACIÓN DE INFORMACIÓN.....	2
TABLA 2. FICHA B 2 RECUPERACIÓN DE INFORMACIÓN.....	4
TABLA 3. FICHA B 3 RECUPERACIÓN DE INFORMACIÓN.....	5
TABLA 4. FICHA B 4 RECUPERACIÓN DE INFORMACIÓN.....	6
TABLA 5. FICHA B 1 WEB SEMÁNTICA.....	8
TABLA 6. FICHA B 2 WEB SEMÁNTICA.....	9
TABLA 7. FICHA B 3 WEB SEMÁNTICA.....	10
TABLA 8. FICHA B 4 WEB SEMÁNTICA.....	11
TABLA 9. FICHA B 1 RECUPERACIÓN DE INFORMACIÓN Y WEB SEMÁNTICA.....	12
TABLA 10. FICHA B 2 RECUPERACIÓN DE INFORMACIÓN Y WEB SEMÁNTICA.....	12
TABLA 11. FICHA B 3 RECUPERACIÓN DE INFORMACIÓN Y WEB SEMÁNTICA.....	14
TABLA 12. FICHA B 4 RECUPERACIÓN DE INFORMACIÓN Y WEB SEMÁNTICA.....	15
TABLA 13. FICHA B 5 RECUPERACIÓN DE INFORMACIÓN Y WEB SEMÁNTICA.....	17
TABLA 14. CASO DE USO INICIAR SESIÓN.....	33
TABLA 15. CASO DE USO REGISTRAR.....	33
TABLA 16. CASO DE USO REALIZAR CONSULTA.....	34
TABLA 17. CASO DE USO CALIFICAR DOCUMENTO.....	34
TABLA 18. CASO DE USO FINALIZAR SESIÓN.....	34
TABLA 19. PRUEBA DE RENDIMIENTO DEL SISTEMA PARA 6 USUARIOS.....	40
TABLA 20. EVALUACIÓN DE USABILIDAD DE LA APLICACIÓN.....	42
TABLA 21. PORCENTAJE DE RESULTADOS PARA CADA PREGUNTA DE USABILIDAD.....	43
TABLA 22. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 22 - SIN EXPANSIÓN.....	45
TABLA 23. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 22 - OPERADORES OR AND (SIN PU).....	45
TABLA 24. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 22 - OPERADORES OR AND (CON PU).....	46
TABLA 25. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 22 - OPERADOR OR (SIN PU).....	46
TABLA 26. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 22 - OPERADOR OR (CON PU).....	47
TABLA 27. CONSULTA 22 - COMPARACIÓN CURVA PRECISION-RECALL DE LA CONSULTA SIN EXPANSIÓN Y LA EXPANSIÓN DE CONSULTA CON OPERADORES OR Y OPERADORES OR-AND EVALUADAS CON Y SIN PU.....	47
TABLA 28. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 26 - SIN EXPANSIÓN.....	49
TABLA 29. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 26 - OPERADORES OR AND (SIN PU).....	49
TABLA 30. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 26 - OPERADORES OR AND (CON PU).....	50
TABLA 31. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 26 - OPERADOR OR (SIN PU).....	51
TABLA 32. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 26 - OPERADOR OR (CON PU).....	51

TABLA 33. CONSULTA 26 - COMPARACIÓN CURVA PRECISION-RECALL DE LA CONSULTA SIN EXPANSIÓN Y LA EXPANSIÓN DE CONSULTA CON OPERADORES OR Y OPERADORES OR-AND EVALUADAS CON Y SIN PU.	51
TABLA 34. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 28 - SIN EXPANSIÓN	53
TABLA 35. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 28 - OPERADORES OR AND (SIN PU)	53
TABLA 36. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 28 - OPERADORES OR AND (CON PU)	54
TABLA 37. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 28 - OPERADOR OR (SIN PU).....	55
TABLA 38. CÁLCULO DE PRECISIÓN Y RECALL CONSULTA 28 - OPERADOR OR (CON PU).....	55
TABLA 39. CONSULTA 28 - COMPARACIÓN CURVA PRECISION-RECALL DE LA CONSULTA SIN EXPANSIÓN Y LA EXPANSIÓN DE CONSULTA CON OPERADORES OR Y OPERADORES OR-AND EVALUADAS CON Y SIN PU.	56
TABLA 40. PRECISIÓN PROMEDIO DE LOS CINCO FORMATOS DE TEXTO APLICADOS A 15 CONSULTAS DE USUARIO DE LA COLECCIÓN MED.	59
TABLA 41. ÍNDICE MAP DE LOS CINCO FORMATOS DE TEXTO APLICADOS A 15 CONSULTAS DE USUARIO DE LA COLECCIÓN MED.....	59
TABLA 42. CONSULTAS UTILIZADAS PARA CALCULAR LA MEDIDA PRECISION AT K.....	59
TABLA 43. APRECIACIÓN DE LOS OBSERVADORES EN K=10 RESULTADOS (MSEC WEB SEARCH – CONSULTA 1).....	60
TABLA 44. EXACTITUD Y PRECISIÓN EN CADA NIVEL DE K (MSEC WEB SEARCH – CONSULTA 1)	60
TABLA 45. APRECIACIÓN DE LOS OBSERVADORES EN K=10 RESULTADOS (MSEC WEB SEARCH – CONSULTA 2).....	60
TABLA 46. EXACTITUD Y PRECISIÓN EN CADA NIVEL DE K (MSEC WEB SEARCH – CONSULTA 2)	61
TABLA 47. APRECIACIÓN DE LOS OBSERVADORES EN K=10 RESULTADOS (MSEC WEB SEARCH – CONSULTA 3).....	61
TABLA 48. EXACTITUD Y PRECISIÓN EN CADA NIVEL DE K (MSEC WEB SEARCH – CONSULTA 3)	61
TABLA 49. APRECIACIÓN DE LOS OBSERVADORES EN K=10 RESULTADOS (GOOGLE – CONSULTA 1)	62
TABLA 50. EXACTITUD Y PRECISIÓN EN CADA NIVEL DE K (GOOGLE – CONSULTA 1).....	62
TABLA 51. APRECIACIÓN DE LOS OBSERVADORES EN K=10 RESULTADOS (GOOGLE – CONSULTA 2)	62
TABLA 52. EXACTITUD Y PRECISIÓN EN CADA NIVEL DE K (GOOGLE – CONSULTA 2).....	63
TABLA 53. APRECIACIÓN DE LOS OBSERVADORES EN K=10 RESULTADOS (GOOGLE – CONSULTA 3)	63
TABLA 54. EXACTITUD Y PRECISIÓN EN CADA NIVEL DE K (GOOGLE – CONSULTA 3).....	63
TABLA 55. APRECIACIÓN DE LOS OBSERVADORES EN K=10 RESULTADOS (YAHOO! – CONSULTA 1).....	64
TABLA 56. EXACTITUD Y PRECISIÓN EN CADA NIVEL DE K (YAHOO! – CONSULTA 1).....	64
TABLA 57. APRECIACIÓN DE LOS OBSERVADORES EN K=10 RESULTADOS (YAHOO! – CONSULTA 2).....	64
TABLA 58. EXACTITUD Y PRECISIÓN EN CADA NIVEL DE K (YAHOO! – CONSULTA 2).....	64
TABLA 59. APRECIACIÓN DE LOS OBSERVADORES EN K=10 RESULTADOS (YAHOO! – CONSULTA 3).....	65

TABLA 60. EXACTITUD Y PRECISIÓN EN CADA NIVEL DE K (YAHOO! – CONSULTA 3).....	65
TABLA 61. CONSULTAS UTILIZADAS PARA CALCULAR LAS ESTADÍSTICAS KAPPA.....	65
TABLA 62. APRECIACIÓN DE LOS OBSERVADORES EN K=10 RESULTADOS PARA LA CONSULTA 1.....	66
TABLA 63. RESULTADO DEL ÍNDICE KAPPA EN K=10 PARA LA CONSULTA 1.....	66
TABLA 64. APRECIACIÓN DE LOS OBSERVADORES EN K=10 RESULTADOS PARA LA CONSULTA 2.....	66
TABLA 65. RESULTADO DEL ÍNDICE KAPPA EN K=10 PARA LA CONSULTA 2.....	67
TABLA 66. APRECIACIÓN DE LOS OBSERVADORES EN K=10 RESULTADOS PARA LA CONSULTA 3.....	67
TABLA 67. RESULTADO DEL ÍNDICE KAPPA EN K=10 PARA LA CONSULTA 3.....	67
TABLA 68. PRECISIÓN PROMEDIO DE LOS RESULTADOS DEVUELTOS POR GOPUBMED Y MSEC WEB SEARCH A TRAVÉS DE 15 CONSULTAS SOBRE MEDLINE.	68
TABLA 69. PRECISIÓN PROMEDIO DE LOS RESULTADOS DEVUELTOS POR GOPUBMED Y MSEC WEB SEARCH A TRAVÉS DE 15 CONSULTAS SOBRE MEDLINE.	69

ÍNDICE DE FIGURAS

FIGURA 1. MODELO DE EXPANSIÓN DE CONSULTA SEMÁNTICA BASADO EN ONTOLOGÍAS.	7
FIGURA 2. DIAGRAMA DE CASOS DE USO.	32
FIGURA 3. VISTA LÓGICA POR CAPAS DE LA APLICACIÓN.	35
FIGURA 4. DIAGRAMA DE DESPLIEGUE.	36
FIGURA 5. VISTA DE IMPLEMENTACIÓN.	37
FIGURA 6. CURVAS PRECISIÓN-RECUERDO PARA LA PRIMERA CONSULTA.	48
FIGURA 7. CURVAS PRECISIÓN-RECUERDO PARA LA SEGUNDA CONSULTA.	52
FIGURA 8. CURVAS PRECISIÓN-RECUERDO PARA LA TERCERA CONSULTA.	56
FIGURA 9. INTERFAZ DE INICIO DE SESIÓN.	70
FIGURA 10. INTERFAZ REGISTRAR USUARIO.	71
FIGURA 11. INTERFAZ DE BÚSQUEDA.	71
FIGURA 12. RESULTADOS DE BÚSQUEDA.	72
FIGURA 13. AYUDA DE LA APLICACIÓN.	73
FIGURA 14. CERRAR SESIÓN.	73

1 ANEXO A

A continuación se presentan las fichas bibliográficas pertenecientes a los núcleos temáticos definidos para este proyecto, los núcleos temáticos fundamentales son la Recuperación de Información y la Web Semántica.

1.1 Fichas bibliográficas Recuperación de Información

Ficha Bibliográfica	
Aspectos formales sobre el documento	
Autor : Miguel A. Alonso, Jorge Graña, Jesús Vilares	
Título: Recuperación de Información en Internet	
Tipo de material: Ponencia	
Enfoque	
Disciplina: señala el área del saber desde donde se define y aborda el objeto de estudio	Recuperación de Información en internet: Principios de la recuperación
Paradigma conceptual: indica la postura teórica y/o metodológica que orienta la investigación.	Estudio de medidas y modelos de recuperación de información
Referentes teóricos: describe los autores específicos en los cuales se apoya el autor y los fundamentos disciplinarios y teóricos que apuntalan la investigación	En varios conceptos del documento, especialmente sobre las cadenas léxicas y la estructura de cohesión léxica, se refieren al autor Morris y sus trabajos de 1991.
Conceptos principales: señala el soporte técnico de la tesis, explicaciones, problemas, ideas y conclusiones planteadas para la investigación	<ul style="list-style-type: none"> • La Recuperación de información (RI, Information Retrieval), es un área de la ciencia y la tecnología que trata de la representación, almacenamiento, organización y acceso a elementos de información. • Un sistema de RI trata de determinar el grado de semejanza de cada uno de los documentos disponibles con la consulta creada por el usuario, para lo cual es necesario discernir entre documentos relevantes y no relevantes de acuerdo a la consulta. • Medidas que dependen de la ordenación de los documentos devueltos. Precisión, R-precisión, Precisión media no interpolada, Precisión media interpolada en 11 puntos. • Los modelos de recuperación de información definen la manera en que se representan las consultas y los documentos.
Hipótesis: registra las proposiciones que sirven de guía a la investigación (conjeturas sobre el objeto de estudio) (suposiciones a partir de los cuales se organizó la investigación)	Su objetivo es presentar el enfoque de la recuperación de información de acuerdo a los modelos y sus medidas
Tesis: describe las proposiciones que sintetizan las generalizaciones sobre el objeto de estudio ("generalizaciones empíricas" existentes en las cuales se apoyó el trabajo)	<ul style="list-style-type: none"> • Un proceso de recuperación de información produce como salida un conjunto de documentos cuyo contenido debe satisfacer la necesidad de información del usuario. • Los sistemas realizan una indexación previa de los documentos relevantes ordenada por el grado de relevancia. A pesar de la eficiencia de muchos de estos sistemas, es difícil unificar la terminología de todo tipo de personas independientemente de su hábitat. • Existen algunas medidas de rendimiento para

	<p>calificar la relevancia de los documentos mostrados al usuario, como la precisión y la cobertura.</p> <ul style="list-style-type: none"> • Se emplean diferentes modelos en los sistemas de recuperación de información, con el fin de mejorar la relevancia de los documentos en las consultas. Los principales son los Booleano y vectorial. Otros también usados son los probabilísticos y basados en semántica latente.
Tipo de investigación: señala el tipo de investigación (exploratoria: examina un tema o problema poco estudiado; Descriptiva: responde al qué del fenómeno analizado; o Explicativa: responde al porqué del fenómeno analizado)	Exploratoria
Metodología	
Tipo de metodología: Describe el tipo de metodología utilizada (Cualitativa: Privilegia la comprensión de los fenómenos y la interacción investigador - investigado; Cuantitativa: Privilegia la cuantificación de los fenómenos estudiados y la no interacción investigador - investigado; o Mixta: Combina procedimientos cualitativos y cuantitativos)	Cualitativa
Técnicas: Describe el tipo de herramientas utilizadas en la recolección, registro y sistematización de la información por el autor en la investigación	Estudios previos de modelos de recuperación de información y de medidas de rendimiento
Resultados:	-
Conclusiones: Señala las conclusiones que especifica explícitamente el autor en el documento	En esta presentación se explora de manera un poco profunda las medidas de precisión y modelos de recuperación de información. Esto es útil para los usuarios que deseen conocer las definiciones y lo más relevante en el área. Es de fácil entendimiento y organización para escudriñar en el tema de la recuperación de información
Recomendaciones: Registra las recomendaciones que hace el autor en el documento	-
Análisis del documento	
Resumen del documento:	Su objetivo es presentar el enfoque de la recuperación de información de acuerdo a los modelos y sus medidas
Palabras claves:	Recuperación de Información
Anexos Referencia los anexos que se consideran relevantes	-
Glosas Indica las inconsistencias, omisiones o errores detectados por quien analizó el documento	-
Comentarios Contiene las reflexiones, relaciones, inferencias y asociaciones que hace quien analizó el documento. También deben indicarse los aportes, vacíos y limitaciones del documento estudiado en relación con el tema central respectivo y/o el área temática correspondiente.	-

Tabla 1. Ficha B 1 Recuperación de Información.

Ficha Bibliográfica
Aspectos formales sobre el documento
Autor: S. Pierre, C. Kacan, and W. Probst
Título: An agent-based approach for integrating user profile into a knowledge management process

Tipo de material: Artículo	
Enfoque	
Disciplina	Recuperación de Información
Paradigma conceptual	Sistemas Basados en Conocimiento
Referentes teóricos	-
Conceptos principales	<ul style="list-style-type: none"> • Perfil de usuario: consiste en un conjunto de especificaciones y características el cual describe un usuario particular. • spherical model: Modelo utilizada para calcular distancia Euclidean entre el vector de peticiones y el vector de documentos para encontrar la similitud entre los términos de solicitud y la base del agente
Hipótesis	<p>En Internet, la información se deriva de los sistemas de información heterogéneos y se presentan bajo una variedad de formatos de archivo. Varias herramientas de búsqueda están siendo utilizados para encontrar información (Yahoo, Alta Vista, Excite, etc.). Algunos de los cuales son más eficientes que otros.</p> <p>Sin embargo, algunos de ellos no permiten a los usuarios reutilizar los resultados obtenidos con el fin de mejorarlas.</p> <p>Otros no tienen en cuenta los intereses o los perfiles de los usuarios en el momento de la búsqueda de información.</p>
Tesis	Para ser eficaz, una herramienta de investigación debe ser capaz de filtrar la información y perfeccionar la búsqueda de acuerdo con el perfil del usuario.
Tipo de investigación	Exploratoria
Metodología	
Tipo de metodología	Mixta
Técnicas	<p>Proponen un modelo de búsqueda híbrida llamado Spherical Model: Que consta de una serie de pasos en que el objetivo es crear documentos de un dominio en particular contenidos en el UA y este extrae sus términos, eliminar términos redundantes, calcula los pesos a cada termino por documento, y calculados los términos el UA permite 2 tipos de búsqueda:</p> <ul style="list-style-type: none"> • Búsqueda conocida por el agente: listado de solicitudes identificados respecto a la solicitud guardada por el usuario • Búsqueda desde una consulta nueva: se crea un vector de pesos a partir de los términos de la consulta y los términos del UA ya analizados sin redundancia y calcula la su similitud utilizando el Spherical Model, luego el usuario evalúa los documentos recuperados entre entre los valores -5 y +5 y de acuerdo a esta evaluación será de gran ayuda en el filtrado y el perfil de usuario será creado y actualizado basándose en aquellas evaluaciones. <p>El enfoque híbrido propuesto en esta sección consiste en combinar los enfoques descritos anteriormente en una forma que permite a un agente de usuario (UA) para llevar a cabo la tarea de búsqueda con una calidad de servicio correspondiente a las necesidades del usuario. El concepto de calidad de servicio se refiere principalmente a la calidad de los resultados y el tiempo de respuesta requerido.</p>
Resultados	-
Conclusiones	Este trabajo pretende un enfoque basado sobre un AGENTE DE USUARIO que permita un número de usuarios conectados en distintas maquinas para acceder a diferentes recursos de información en un orden que satisfaga la solicitud de búsqueda.
Recomendaciones	La base para el nuevo enfoque consiste en permitir la utilización de la información sobre el perfil para refinar los resultados proporcionados por los modelos descritos en este trabajo.
Análisis del documento	
Resumen del documento	La enorme cantidad de información disponible induce el problema de la sobrecarga de la información, hay demasiada información para que los usuarios la pueda digerir o asumir. Para solucionar este problema se propone una arquitectura inteligente/interactiva IIR agente de recuperación de información basada en agentes para integrar varias funciones en la IR de la web.

Palabras claves	Recuperación de Información, Perfiles de Usuario, Sistemas Multiagente
Anexos	-
Glosas	-
Comentarios	No se plantea una arquitectura, no se ha probado el agente (UA)

Tabla 2. Ficha B 2 Recuperación de Información.

Ficha Bibliográfica	
Aspectos formales sobre el documento	
Autor: Pei-Min Chen and Fong-Chou Kuo	
Título: An information retrieval system based on a user profile	
Tipo de material: Artículo	
Enfoque	
Disciplina	Recuperación de Información
Paradigma conceptual	Sistemas Basados en Conocimiento
Referentes teóricos	-
Conceptos principales	Búsqueda personalizada
Hipótesis	Debido a las características de los métodos de recuperación, los modelos de IR convencionales sufren a menudo de consultas inexactas e incompletas, así como la pertinencia de documentos inconsistentes. La exactitud es difícil de lograr entonces se intenta es mejorarla.
Tesis	-
Tipo de investigación	Exploratoria
Metodología	
Tipo de metodología	Mixta
Técnicas	Query enhancement. Este subproceso extrae la keywords denominados términos de consulta desde la consulta editada. También identifica términos SR o CO que correspondan a los términos de la consulta. Si sus grados de SR o CO son mayores que el valor predefinido 0.5. Aquellos 3 clases de términos son combinados con operadores lógicos, AND o OR para formar una consulta mejorada entonces los términos de la consulta que sean sr son OR y los CO son AND juntos.
Resultados	También se muestra que la recuperación / tasas de precisión del sistema son aproximadamente el 91% y 81%, respectivamente, en promedio.
Conclusiones	Este proyecto que se acerca cada vez más o lo que el usuario necesita de una consulta por medio de una arquitectura que busca retroalimentarse de acuerdo a lo que el usuario especifique en su consulta y poder predecir mas cual es el perfil del usuario.
Recomendaciones	-
Análisis del documento	
Resumen del documento	Este trabajo propone mejorar la precisión de la IR mediante el diseño de un modelo de IR que puede adaptarse a las diversas necesidades de un individuo y puede realizar una búsqueda personalizada. En este trabajo, una adaptación del sistema de recuperación de información integrada en un mecanismo de retroalimentación inteligente de ajuste, se propone captar las nociones personales de los términos de la consulta. Este mecanismo, junto con una tabla de correspondencias que se utiliza como un perfil de usuario para simular las nociones del usuario de los términos. Esta tabla de correspondencias que incluye los grados de relevancia semántica (SR) y co-ocurrencia (CO) entre los términos. Al usar el perfil personal durante la recuperación, el sistema puede obtener una mejor correspondencia entre las necesidades de información y los resultados recuperados.
Palabras claves	Recuperación de Información, Perfil de Usuario
Anexos	-
Glosas	No se habla de proyectos a futuro o si ya se a puesto en práctica esta arquitectura ya que solo han hecho un experimento no más.
Comentarios	La consulta proporcionada del usuario se modifica si se sobrepasa de un valor predefinido por el sistema, modificando dicha consulta para una mejor IR.

Tabla 3. Ficha B 3 Recuperación de Información.

Ficha Bibliográfica	
Aspectos formales sobre el documento	
Autor: Vadim Paz Madrid Gorelov, Ángel F. Zazo, Carlos G. Figuerola, José Luis Alonso Berrocal	
Título: Librerías Lucene y dotLucene para Recuperación de Información. Estudio y desarrollo de casos prácticos	
Tipo de material: Informe técnico	
Enfoque	
Disciplina	Tutorial básico de la utilización de la librería Lucene.
Paradigma conceptual	Se estudia el modelo de clases de Lucene y se explotan los principales objetos para la indexación y la búsqueda de información. Además se estudia la librería Lucene y se describe la utilización de dotLucene.
Referentes teóricos	Greenstette: dice que el proceso de separación de palabras se compone de un montón de preguntas espinosas, de las cuales solamente unas pocas tienen respuesta perfecta.
Conceptos principales	<ul style="list-style-type: none"> • Término índice: es una palabra que posee significado y que se utiliza para representar un concepto. • Los sistemas de información no solo persiguen encontrar aquellos términos que mejor representen a los documentos, sino además aquellos que permiten diferenciar unos respecto a otros. Los diferentes modelos de recuperación se diferencian en el método elegido para representar los documentos y las consultas y para realizar las búsquedas de información. • Análisis léxico del texto, con el objetivo de convertir la cadena de entrada en un conjunto de palabras, y determinar el tratamiento que se realizará sobre números, signos de puntuación, tratamiento de mayúsculas y/o minúsculas, nombres propios, etc. Los separadores de palabras en español suelen ser el espacio y un conjunto más o menos reducido de signos de puntuación. • Eliminación de palabras vacías, muy frecuentes y muy poco frecuentes, con el objetivo de reducir el número de términos con valores muy poco significativos para la recuperación. Es también una buena forma de reducir el tamaño de los ficheros de datos del sistema, y así aumentar la rapidez de respuesta del sistema. • Aplicación de lematización sobre los términos resultantes para eliminar variaciones morfosintácticas y obtener lemas. La lematización consiste en elegir convencionalmente una forma de una palabra para remitir a ella todas las de su misma familia por razones de economía. • Selección de términos o grupos de términos que serán considerados términos índice. Normalmente se realiza sobre la naturaleza morfo-sintáctica del término, pues, como ya se ha mencionado, los términos que actúan gramaticalmente como nombres suelen poseer un mayor contenido semántico que verbos, adjetivos o adverbios.
Hipótesis	Los sistemas de información son la pieza clave para resolver el problema de la búsqueda rápida y eficiente de la información, por lo cual hoy en día se debe tener la capacidad de encontrar información y datos de diversos tipos y formatos de manera flexible, libre de forma que permita realizar búsquedas requiriendo el mínimo esfuerzo posible.
Tesis	Se realiza una introducción a la apasionante disciplina que es la recuperación de información y la utilización de una de las más flexibles y poderosas librerías del mercado para crear aplicaciones de recuperación de información denominada Lucene.
Tipo de investigación	Descriptiva
Metodología	
Tipo de metodología	Mixta
Técnicas	Lucene, dotLucene
Resultados	-
Conclusiones	Como resultado de su desarrollo se pudo concluir lo siguiente: <ul style="list-style-type: none"> • Tanto Lucene como dotLucene son librerías fácilmente reutilizables

	<p>para aplicaciones que requieran agregar la capacidad de búsqueda en aplicaciones que contengan una fuente de datos definida y su información se obtenga como texto.</p> <ul style="list-style-type: none"> • Una de las mejores ventajas de Lucene es que al ser un proyecto de código abierto y de rápida implementación, ofrece la posibilidad de agregar nuevas funcionalidades desarrollando nuevos componentes. En nuestro caso hemos agregado un analizador léxico que reconoce los patrones propios de la Lengua Española, permitiendo de esta manera la búsqueda de términos en este idioma. • Las aplicaciones creadas deben considerarse puntos de partida para el estudio y ampliación de la potencia de Lucene como herramienta de recuperación de información, fortalecida fuertemente por su flexibilidad en la obtención y asignación de parámetros de configuración iniciales, lo que las hace portables y demostrables hacia diferentes ambientes y fuentes de datos que sigan las pautas de los ejemplos mostrados en nuestro trabajo. • Finalmente, se resalta que Lucene define un modelo de clases compacto y de fácil comprensión, por lo que una implementación inicial completa de búsqueda e indexación se puede realizar con muy pocas líneas de código y pocas instancias de objetos de Lucene. A partir de este modelo se pueden ir agregando funcionalidades basándose en la exploración del modelo completo de clases de Lucene.
Recomendaciones	-
Análisis del documento	
Resumen del documento	<p>En este informe técnico se describe la utilización de dos librerías¹ para Recuperación de Información. Después de una introducción a esta disciplina, se realiza un tutorial básico de utilización de la librería Lucene, bajo el lenguaje de programación Java, explicando en qué consiste, qué se puede hacer con ella, y poniendo ejemplos prácticos de su utilización. Se estudia el modelo de clases de Lucene, y se exploran los principales objetos para la indexación y búsqueda de información. Además del estudio e implementación de la librería Lucene, se describe la utilización de dotLucene, un puerto adicional de Lucene en .Net, con el que probar la versatilidad de Lucene en otras plataformas. Para ello se han elaborado y documentado dos ejemplos de búsqueda de información. En el primero se lleva a cabo una búsqueda de información en documentos almacenados en un árbol de directorios. Se pueden realizar búsquedas de información sobre cualquier fichero convertible a texto plano. El segundo va más allá y realiza la indexación, delimitación y búsqueda de información en documentos XML, permitiendo la búsqueda por campos concretos en este tipo de documentos.</p>
Palabras claves	Lucene, recuperación de información
Anexos	-
Glosas	-
Comentarios	-

Tabla 4. Ficha B 4 Recuperación de Información.

1.2 Fichas bibliográficas Web Semántica

Ficha Bibliográfica	
Aspectos formales sobre el documento	
Autor: Hongsheng Wang, Jiuying Qin and Hong Shao	
Título: Expansion Model of Semantic Query Based on Ontology (Modelo de expansión de Consulta Semántica Basado en Ontologías)	
Tipo de material: Artículo	
Enfoque	
Disciplina	Recuperación de Información
Paradigma conceptual	Ingeniería del Conocimiento
Referentes teóricos	<ul style="list-style-type: none"> • XU J X, CROFT W B. Análisis Global de Documentos.

	<ul style="list-style-type: none"> • DEERWESTER S, DUMAI ST, FURNAS G W, et al. Algoritmos de Clustering e Índices Semánticos Latentes. • QIU Y, FREI H. Algoritmo de un Diccionario de Similitud.
Conceptos principales	<p>Expansión Semántica de la Consulta.</p> <p>Diferencia de Expresión: Se refiere a que un término de la consulta puede tener muchos otros términos equivalentes, es decir que representen un mismo concepto.</p> <p>Emparejamiento Mecánico: Usualmente los tradicionales modelos de RI sólo recuperan documentos que contengan el término exacto, dejando de recuperar documentos que involucren el mismo concepto pero representado en diferentes términos.</p>
Hipótesis	<p>Un sistema de RI puede mejorar la semántica entendiendo los documentos y al mismo tiempo brindando la oportunidad al usuario de establecer su necesidad de información de una forma más natural (más precisa), lo cual seguramente optimizará todo el proceso de búsqueda y aumentará la relación de las palabras de la consulta y los documentos, además de mejorar la precisión y el recuerdo de la RI.</p>
Tesis	<p>Atacar dos de los principales problemas de la RI, tales como:</p> <ul style="list-style-type: none"> • Diferencia de Expresión • Emparejamiento Mecánico <p>Se propone un modelo compuesto por cuatro módulos:</p> <ul style="list-style-type: none"> • Módulo de Interfaz de Usuario • Módulo manejador de consulta • Módulo de Recuperación • Módulo manejador de resultados <p>En la Figura 1 se presenta una arquitectura del modelo:</p> <p>Figura 1. Modelo de expansión de Consulta Semántica Basado en Ontologías.</p> <p>Para la expansión se extraen conceptos relacionados como sinónimos, hiperónimos, hipónimos y conceptos hermanos; a los cuales se calcula el grado de relación. Aquellos más relacionados se consideran como palabras de expansión.</p>
Tipo de investigación	Exploratoria
Metodología	
Tipo de metodología	Tipo de Metodología: Mixta (Cualitativa y Cuantitativa)
Técnicas	<ul style="list-style-type: none"> • Se construyó una Ontología y se introdujo en el modelo • A partir de 10 consultas se calculan los radios de precisión y recuerdo; y se comparan con los métodos tradicionales de búsqueda.
Resultados	-
Conclusiones	En la experimentación se probó adicionando de 3 a 18 palabras de expansión, los mejores resultados se obtuvieron adicionando 6 palabras.

	Con más de 6 palabras, la precisión decrece porque demasiadas palabras añaden "ruido" al proceso de búsqueda.
Recomendaciones	-
Análisis del documento	
Resumen del documento	Un modelo de expansión de consulta semántica basada en ontologías, se propone para el problema de la "Diferencia de Expresión" y "Coincidencia Mecánica" que por lo general existe en la recuperación de la información. Mediante el uso de conocimiento del dominio en la ontología, el sistema de recuperación puede mejorar la comprensión semántica de los documentos recuperados, y dar la oportunidad al usuario a enviar su solicitud de información en forma más natural (una forma más precisa), que sin duda optimizará todo el proceso de búsqueda y aumentará la relevancia de palabras de consulta y documentos, además de mejorar las relaciones de recuperación y precisión de recuperación de información. Para llevar a cabo la expansión semántica basada en ontologías, el algoritmo de expansión semántica y las funciones de expansión semántica basada en ontologías son propuestos y utilizados en el modelo. Como se indica en el experimento, este modelo es adecuado para promover la recuperación de información en línea y la precisión de las búsquedas es mejorada.
Palabras claves	Expansión de Consulta, Ontologías
Anexos	-
Glosas	-
Comentarios	Falencias: <ul style="list-style-type: none"> • No se hace uso del Perfil de Usuario en el proceso de Expansión. • Para el proceso de expansión sólo hace uso de la propia Ontología de dominio.

Tabla 5. Ficha B 1 Web Semántica.

Ficha Bibliográfica	
Aspectos formales sobre el documento	
Autor: Roberto Navigli and Paola Velardi	
Título: An Analysis of Ontology-based Query Expansion Strategies (Un análisis de las estrategias de expansión de consultas basado en ontologías)	
Tipo de material: Artículo	
Enfoque	
Disciplina	Recuperación de Información
Paradigma conceptual	Ingeniería del Conocimiento
Referentes teóricos	<ul style="list-style-type: none"> • Cutker et al. Análisis de Estructura de Documentos. • Carpineto et al. Expansión de consulta usando métodos estadísticos. • Cui et al. Logs de consulta. • Gonzalo et al. Efecto de la expansión de consulta con synsets de WordNet.
Conceptos principales	Expansión de consulta basada en significado, Word Sense Disambiguation (WSD).
Hipótesis	<ul style="list-style-type: none"> • La expansión con sinónimos e hiperónimos tiene un efecto limitado en la RI, mientras que otros tipos de información semántica derivables desde una Ontología son mucho más efectivos para mejorar los resultados de búsqueda. • Se parte de la premisa de que los resultados van a ser mejores si se utilizan los motores de búsqueda más populares.
Tesis	<ul style="list-style-type: none"> • Este estudio se centra en cómo usar la información del significado (y las Ontologías en general) para expandir la consulta. • Se prueban 5 métodos de expansión: Sinónimos, Hiperónimos, Synset de la definición, palabras de la definición, por nodos comunes (se encuentran en el conjunto S_j). • En WordNet existe una confusión Ontológica entre instancias y conceptos, estos son tratados como el mismo objeto.

Tipo de investigación	Exploratoria
Metodología	
Tipo de metodología	Tipo de Metodología: Mixta (Cualitativa y Cuantitativa)
Técnicas	Para evaluar el algoritmo de desambigüedad de significado de palabra o sus siglas en inglés WSD, se seleccionan 5 métodos de expansión y dos estrategias de selección de palabras.
Resultados	-
Conclusiones	<ul style="list-style-type: none"> • La expansión con sinónimos tiene un efecto limitado sobre la RI. • La expansión con hiperónimos es la estrategia que ofrece los peores resultados de acuerdo a los 5 métodos estudiados, por ende, los nodos más generales causan ruido en la expansión de consulta. • La técnica de "nodos comunes" produce buenos resultados sólo cuando los nodos comunes seleccionados están relacionados en una relación No Taxonómica. • En cuanto a resultados se observa que el método de expansión por "palabras de la definición", logra mejores resultados que los demás.
Recomendaciones	
Análisis del documento	
Resumen del documento	La expansión consulta basada en el sentido nunca ha demostrado su eficacia a excepción de la tarea llamada "respuesta de pregunta de dominio abierto". El presente trabajo sigue sin ser concluyente en este sentido, debido a algunas limitaciones experimentales, pero se aportan pruebas interesantes que sugieren las nuevas directrices para futuras investigaciones. La desambiguación de sentido de la palabra es, de hecho, sólo uno de los problemas que implica la expansión de consultas basada en sentido. El segundo es el uso de información sensorial (y ontologías en general) para ampliar la consulta. Se demuestra que la expansión con sinónimos o hiperónimos tiene un efecto limitado sobre el rendimiento de recuperación de información web, mientras que otros tipos de derivar la información semántica de una ontología son mucho más eficaces en la mejora de resultados de búsqueda.
Palabras claves	Ontologías, Expansión de Consulta, WSD
Anexos	-
Glosas	-
Comentarios	<p>Falencias:</p> <ul style="list-style-type: none"> • Existe una limitación de Google a 10 palabras por consulta, razón por la cual no es la mejor elección para probar el algoritmo propuesto. • Para consultas mas largas existe el problema de seleccionar solo una porción de las palabras de expansión, en este caso se escogen las primeras palabras de las listas, esta estrategia se optimiza en el método de expansión por nodos comunes puesto que el conjunto S_x se encuentra ordenado de acuerdo a la fortaleza de la intersección de cada synset S.

Tabla 6. Ficha B 2 Web Semántica.

Ficha Bibliográfica	
Aspectos formales sobre el documento	
Autor: Ana G. Maguitman, Filippo Menczer, Fulya Erdinc, Heather Roinestad and Alessandro Vespignani	
Título: Algorithmic Computation and Approximation of Semantic Similarity (Cálculo Algorítmico y aproximación de Similitud Semántica)	
Tipo de material: Artículo	
Enfoque	
Disciplina	Ciencias de la Computación
Paradigma conceptual	Ingeniería del Conocimiento
Referentes teóricos	-
Conceptos principales	Similitud Semántica
Hipótesis	Cómo estimar la similitud semántica en Ontologías generalizadas, tomando las ventajas tanto de los componentes jerárquicos como de los no jerárquicos.

Tesis	<p>No es apropiado utilizar muchas relaciones No Jerárquicas.</p> <ul style="list-style-type: none"> • Permitir múltiples enlaces cruzados es a nivel computacional inviable porque esto conduce a una membresía de temas muy densa, por ejemplo todos los temas pertenecen a casi cualquier otro tema. • Los enlaces poco confiables producen cambios muy significativos en las funciones de membresía. • Cada camino haría la clasificación por significado mezclando todos o una gran cantidad de temas <p>Se utiliza máximo una relación No Jerárquica</p> <ul style="list-style-type: none"> • Permite capturar los componentes no-jerárquicos de la ontología, preservando la viabilidad, robustez y significado <p>Se obtiene una métrica para la Similitud Semántica de forma general a partir de la siguiente fórmula:</p> $\sigma_s^G(t_1, t_2) = \max_k \frac{2 \cdot \min(W_{k1}, W_{k2}) \cdot \log Pr[t_k]}{\log(Pr[t_1 t_k] \cdot Pr[t_k]) + \log(Pr[t_2 t_k] \cdot Pr[t_k])}$
Tipo de investigación	Exploratoria
Metodología	
Tipo de metodología	Tipo de Metodología: Mixta (Cualitativa y Cuantitativa)
Técnicas	Teniendo en cuenta máximo un enlace a través de cada ruta de pertenencia, permite capturar los componentes no-jerárquicos de la ontología, preservando la viabilidad, robustez y significado.
Resultados	-
Conclusiones	<ul style="list-style-type: none"> • En este trabajo se presenta una nueva medida de similitud semántica de las páginas Web que generaliza la bien fundada medida la teoría de la información semántica de árboles basado en similitud con el caso en que las páginas se clasifican en los nodos de una ontología gráfica arbitraria, en componentes tanto jerárquicos como no jerárquicos. Esta medida puede aplicarse fácilmente a los datos de la semántica de ontologías de actualidad y directorios Web, como Yahoo, PAO y sus derivados. • La metodología descrita para evaluar los algoritmos de clasificación basados en la similitud semántica se puede aplicar a combinaciones arbitrarias de las funciones de clasificación derivadas de análisis de textos, análisis de enlaces, y otras características disponibles para un motor de búsqueda.
Recomendaciones	-
Análisis del documento	
Resumen del documento	<p>La extracción automática de información semántica del texto y los enlaces en las páginas Web es clave para mejorar la calidad de los resultados de búsqueda. Sin embargo, la evaluación automática de las medidas semántica es limitada por el alcance de los estudios de usuarios, que no escala con el tamaño, la heterogeneidad, y el crecimiento de la Web. Este artículo se propone aprovechar los metadatos generados por el hombre para medir las relaciones semánticas entre un número masivo de pares de páginas Web o temas.</p> <p>Este documento define una medida de información de la teoría de la similitud semántica que explota tanto la estructura jerárquica y no jerárquica de una ontología. Un estudio experimental muestra que esta medida mejora de manera significativa en el enfoque tradicional basado en la taxonomía.</p>
Palabras claves	Ontologías, Similitud Semántica
Anexos	-
Glosas	-
Comentarios	-

Tabla 7. Ficha B 3 Web Semántica.

Ficha Bibliográfica
Aspectos formales sobre el documento

Autor : Shahrul Azman Noah, Lailatulqadri Zakaria, Arifah Che Alhadi, Tengku Mohd Tengku Sembok, Saidah Saad	
Título: Towards Building Semantic Rich Model for Web Documents Using Domain Ontology (Hacia la construcción de un modelo rico en semántica para documentos web usando ontologías de dominio)	
Tipo de material: Artículo	
Enfoque	
Disciplina	Área de la informática
Paradigma conceptual	Web Semántica
Referentes teóricos	-
Conceptos principales	Extracción de información semántica contenida en documentos web.
Hipótesis	Se desarrolla una herramienta para extraer la información semántica contenida en los documentos web, mediante el análisis del lenguaje natural y la ontología de dominio específico. La idea es construir una base de datos específica que contenga los índices semánticos de los documentos. Esto está destinado a servir e integrar documentos de sistemas de recuperación semántica dentro de un grupo o comunidades.
Tesis	-
Tipo de investigación	Exploratoria
Metodología	
Tipo de metodología	Descriptiva
Técnicas	-
Resultados	-
Conclusiones	La ontología de dominio desempeña un papel importante en el apoyo a las tareas de clasificación de documentos y la organización de estos mismos. En este trabajo se presento una ontología de dominio combinado con una técnica de análisis del lenguaje natural, que puede ser explotado no solo para extraer los conceptos más importantes de los documentos sino también para construir el contenido semántico de los documentos web.
Recomendaciones	-
Análisis del documento	
Resumen del documento	El acceso y la extracción de significados semánticos de documentos web son cruciales para la realización de la Web Semántica. Si bien la web ofrece la flexibilidad de hacer la información fácilmente disponible, es bastante difícil encontrar un camino de describir, clasificar y presentar esta información con contenido semántico ricos. Por lo tanto, la semántica contenido de la información de los documentos web deben ser determinados con el fin de hacer la información más accesible a los motores de búsqueda y otras aplicaciones. En este trabajo se propone un enfoque destinado a ayudar en la construcción de modelos de documentos semánticos con técnica de análisis de lenguaje natural y un dominio específico de la ontología. .
Palabras claves	Web Semántica, Extracción de Contenido Web
Glosas	-
Comentarios	-

Tabla 8. Ficha B 4 Web Semántica.

1.3 Fichas bibliográficas Recuperación de Información y Web Semántica

Ficha Bibliográfica	
Aspectos formales sobre el documento	
Autor: Song Jun-feng	
Título: Ontology-Based Information Retrieval Model for The Semantic Web (ontología basada en modelo de recuperación de información para la web semántica)	
Tipo de material: Artículo – tesis	
Enfoque	
Disciplina	Indexación Semántica y ontologías de dominio
Paradigma conceptual	Se enfocan en la generación de una ontología teniendo en cuenta la traducción e integración de ontologías de dominio. Además se marca el

	contenido semántico por medio de un razonador lógico de descripción
Referentes teóricos	-
Conceptos principales	•
Hipótesis	•
Tesis	•
Tipo de investigación	Descriptiva
Metodología	
Tipo de metodología	Mixta: combina la descripción teórica y practica mostrando el modelo de generación de la ontología.
Técnicas	Documentos bibliográficos, varias ontologías de dominio, un razonador lógico de descripción. Lenguaje OWL Lite para la ontología
Resultados	-
Conclusiones	
Recomendaciones	
Análisis del documento	
Resumen del documento	-
Palabras claves	Ontologías de dominio, razonador lógico, términos de índices semánticos.
Glosas	-
Comentarios	-

Tabla 9. Ficha B 1 Recuperación de Información y Web Semántica.

Ficha Bibliográfica	
Aspectos formales sobre el documento	
Autor : Gang Lv, Cheng Zheng and Li Zhang	
Título : Text Information Retrieval Based on Concept Semantic Similarity (Recuperación de Información Textual Basada en el Concepto de Similitud Semántica)	
Tipo de material : Artículo - Tesis	
Enfoque	
Disciplina	Recuperación de Información, Similitud Semántica.
Paradigma conceptual	
Referentes teóricos	•
Conceptos principales	
Hipótesis	
Tesis	La idea básica del método para calcular la similitud semántica es cuantificarla con la distancia geométrica de dos conceptos (nodos) en una red.
Tipo de investigación	Descriptiva
Metodología	
Tipo de metodología	Metodología Mixta: combina procedimientos cualitativos mostrando los pasos de la investigación y el proceso; así mismo muestra en valores y porcentajes los resultados del estudio.
Técnicas	Fgfg Hjh
Resultados	
Conclusiones	
Recomendaciones	-
Análisis del documento	
Resumen del documento	
Palabras claves	Recuperación de Información, Ontologías, Similitud Semántica, Expansión de Consulta, Pesos de Documentos
Anexos	-
Glosas	No se mencionan posibles recomendaciones o trabajo futuro.
Comentarios	La ontología utilizada es muy simple por lo que no permite un razonamiento semántico más amplio que permita mejores resultados en la expansión de consulta.

Tabla 10. Ficha B 2 Recuperación de Información y Web Semántica.

Ficha Bibliográfica	
Aspectos formales sobre el documento	
Autor : Marie Aude Aufaure, Rania Soussi, Hajer Baazaoui.	
Título: SIRO: On- Line Semantic Information Retrieval Using Ontologies (Recuperación de información semántica en línea usando ontologías)	
Tipo de material: Artículo	
Enfoque	
Disciplina	Ontologías de dominio, repositorios como WordNet, vector modelo.
Paradigma conceptual	Proponer una solución a la recuperación de la información a través de ontologías ya que estas pueden ayudar al usuario a encontrar documentos de un dominio específico.
Referentes teóricos	-
Conceptos principales	<ul style="list-style-type: none"> • Se recalca el gran crecimiento que ha tenido la Web, y el problema que esto presenta ya que la recuperación de información es cada vez más difícil. • Las ontologías pueden mejorar la llamada y la precisión, eliminando la ambigüedad del lenguaje natural utilizando axiomas y conceptos. • Las ontologías pueden ayudar al usuario a formular sus necesidades y tener acceso a los documentos.
Hipótesis	<ul style="list-style-type: none"> • Proponer un sistema basado en ontologías que de alguna forma ayude a recuperar la información de una manera más eficiente. • Las ontologías son uno de los mecanismos mas utilizados para resolver el problema de la recuperación de la información. • Hacen uso de un dominio de ontología combinado con un servicio de ontología • Hacen una adaptación del vector modelo, sustituyendo los términos por conceptos. • La pregunta del usuario es enriquecida usando conceptos y relaciones del dominio ontológico y WordNet. • Este proyecto es aplicado para dominio turístico. • Este proyecto puede ser aplicado en diferentes campos.
Tesis	<ul style="list-style-type: none"> • El enorme número de documentos disponibles en la web hace difícil encontrar lo que es pertinente. Por lo tanto, la búsqueda de información se vuelve más y más compleja, debido al creciente volumen de datos y de su falta de estructura. La calidad de los resultados que los tradicionales motores de búsqueda de texto proporcionan todavía no es óptima para muchos tipos de consultas de los usuarios. Especialmente, las ambigüedades de los lenguajes naturales y los conceptos abstractos son manejados inadecuadamente por los motores de búsqueda de texto. Ontologías ofrecer una solución a estos problemas. Ellas pueden ayudar a un usuario a encontrar los documentos de un dominio específico. Este documento propone un sistema de recuperación basado en ontologías. Este sistema integra los resultados de los tradicionales motores de texto completo, y por lo tanto apoya una transición gradual desde los clásicos motores de búsqueda de texto basado en las ontologías. • El propósito del proyecto ON-LINE IR es usar dos tipos de ontologías las cuales son ontologías de dominio y ontología de servicio y wordnet. La ontología de dominio que es construida manualmente es la base de la recuperación ON-LINE y contiene un dominio que es caracterizado por una lista de servicios, actividades y tareas.
Tipo de investigación	Exploratoria
Metodología	
Tipo de metodología	Metodología Mixta: combina procedimientos cualitativos mostrando los pasos de la investigación y el proceso; así mismo muestra en valores y porcentajes los resultados del estudio.
Técnicas	Para el proyecto hacen uso del repositorio WordNet, para lematización es usado TreeTagger, para el parseo de las páginas web utilizan DOM, la herramienta para el manejo de las ontologías usan Jena Api y Google Api para buscar en la Web. Para el vector modelo ellos se basan en el modelo propuesto por Salton.

Resultados	-
Conclusiones	Este sistema se ha realizado para aplicarlo al dominio turístico y se señala que permite mejorar la precisión de la investigación y por lo tanto la pertinencia de los documentos devueltos al usuario. Este trabajo da al usuario la oportunidad de detectar los servicios de dominio exacto.
Recomendaciones	SIRO como es llamado este sistema se puede aplicar a cualquier campo, este trabajo también hace que sea posible proporcionar una preclasificación y filtrado de documentos para mejorar la construcción de ontologías basadas en técnicas de aprendizaje. En el futuro, este sistema se integrará a un prototipo desarrollado para la construcción semiautomática de ontologías de dominio y de servicio.
Análisis del documento	
Resumen del documento	El enorme número de documentos disponibles en la Web hace que encontrar los documentos necesarios sea una tarea difícil para usuarios. Por lo tanto, la búsqueda de información se vuelve más y más compleja, debido al creciente volumen de datos y de su falta de estructura. La calidad de los resultados de los tradicionales motores de búsqueda de texto completo no es óptima para muchos tipos de consultas de los usuarios. Especialmente, las ambigüedades de los lenguajes naturales y los conceptos abstractos son manejados inadecuadamente por los motores de búsqueda de texto completo. Las Ontologías ofrecen una solución a estos problemas. Ellas pueden ayudar a un usuario a encontrar los documentos de un dominio específico. Este proyecto propone un sistema de recuperación basado en ontologías. Este sistema integra los resultados de los tradicionales motores de texto completo, y por lo tanto apoya una transición gradual desde el clásico motor de búsqueda de texto completo a la ontología.
Palabras claves	Ontologías, Modelo Espacio Vectorial
Anexos	-
Glosas	-
Comentarios	-

Tabla 11. Ficha B 3 Recuperación de Información y Web Semántica.

Ficha Bibliográfica	
Aspectos formales sobre el documento	
Autor: Jibrán Mustafa, Sharifullah Khan, Khalid Latif.	
Título: Ontology Based Semantic Information Retrieval (Ontología basada en la recuperación de información semántica)	
Tipo de material: Artículo	
Enfoque	
Disciplina	Ontologías, RDF triplets.
Paradigma conceptual	Recuperación de Información
Referentes teóricos	-
Conceptos principales	<ul style="list-style-type: none"> • Ontologías de dominio se utiliza como base de conocimiento para entender los significados de los conceptos. • Una ontología es una especificación formal explícita de una conceptualización compartida • Similitud temática interpreta el significado de las palabras claves y sus relaciones, ellas utilizan una temática similar emparejando en RDF triples para concentrarse en ambos aspectos (conceptos y sus relaciones)
Hipótesis	<p>Para este trabajo se propone una temática similar a otros trabajos, en donde en vez de utilizar palabras claves se emplea RDF triples y se concentra en el contexto de la búsqueda del término. Para tal propósito se utiliza los siguientes componentes: crawler, source model, semantic matcher, query reformulator and ranker.</p> <p>El sujeto, la propiedad y el objeto de RDF triples permite al armazón de la búsqueda concentrarse en la combinación de conceptos y su relación de similitud al mismo tiempo.</p> <p>Algunos de los componentes a utilizar:</p>

	<p>Crawler: extrae metadatos en la forma RDF triples desde documentos residentes en un repositorio de información y los carga en un modelo fuente.</p> <p>Modelo fuente: este contiene dos componentes:</p> <ul style="list-style-type: none"> • Metadato de origen: contiene metadatos contenidos en documentos de la forma RDF triples. • Metadato de contenido: ayuda a la identificación de documentos relevantes. <p>Semantic macher:</p> <p>Re formulador de pregunta: La pregunta del usuario es reformulada utilizando sinónimos y otras relaciones, entonces la pregunta se vuelve a escribir utilizando los términos mencionados y esta pregunta se pasa al matcher semántico en la forma de RDF triples.</p>
Tesis	Se propone un marco de recuperación de información semántica para mejorar la precisión de los resultados de las búsquedas y se emplea el enfoque de la similitud temática para la recuperación de información, a fin de captar el contexto del concepto particular.
Tipo de investigación	Exploratoria
Metodología	
Tipo de metodología	Metodología Mixta: combina procedimientos cualitativos mostrando los pasos de la investigación y el proceso; así mismo muestra en valores y porcentajes los resultados del estudio.
Técnicas	Resource Description Framework (RDF) triplets.
Resultados	Los experimentos realizados en este proyecto arrojaron excelentes resultados.
Conclusiones	Los resultados de los experimentos realizados a este proyecto han demostrado tener unos excelentes resultados ha comparación de otras técnicas utilizadas para la recuperación de la información, lo que hace que los usuarios obtengan unos resultados mas precisos en sus búsquedas.
Recomendaciones	Un trabajo futuro se pretende aumentar el armazón para otros datos y tratar con otros datos heterogéneos.
Análisis del documento	
Resumen del documento	Semántica basada en las técnicas de recuperación de información comprende los significados de los conceptos que los usuarios especifiquen en sus consultas. El principal inconveniente de la semántica existente basado en técnicas de recuperación de información es que ninguno de ellos considera el contexto del concepto (s). Se propone un marco de recuperación de información semántica para mejorar la precisión de los resultados de búsqueda. En este trabajo, el enfoque de la similitud temática se emplea para la recuperación de información, a fin de captar el contexto del concepto particular (s). Ellos guardan la información de metadatos de fuente (s) en forma de triples RDF. Buscan las preguntas de los usuarios en los metadatos existentes, haciendo coincidir triples RDF en lugar de palabras clave. Los resultados de los experimentos realizados en el marco mostró mejoras en la precisión y el recuerdo en comparación con la semántica existente basado en técnicas de recuperación de información.
Palabras claves	Semántica, recuperación de información, temática similar, metadatos, RDF.
Anexos	-
Glosas	-
Comentarios	-

Tabla 12. Ficha B 4 Recuperación de Información y Web Semántica.

Ficha Bibliográfica
Aspectos formales sobre el documento
<p>Autor: Mingxia Gao, Chunnian Liu, Furong Chen</p> <p>Título: An Ontology Search Engine Based on Semantic Analysis (Un motor de búsqueda ontológico basado en el análisis semántico)</p>

Tipo de material: Artículo	
Enfoque	
Disciplina	Ontologías, algoritmo WI-OUTOSEARCH.
Paradigma conceptual	Recuperación de Información
Referentes teóricos	-
Conceptos principales	<ul style="list-style-type: none"> • Una ontología está formada por varios términos (que se denominan conceptos) que están relacionados y limitados por diversas limitaciones estructurales. • TUCUXI captura la semántica de las páginas web a través de herramientas lingüísticas como wordnet y devuelven los resultados adecuados. • Punto de referencia ontología: Dada una Ontología, que es una norma por la que la ontología de otros se puede medir o juzgar. • Evaluación de la ontología: una ontología que debe comparar con la ontología de referencia.
Hipótesis	<ul style="list-style-type: none"> • Se propone un vector de pesos y un algoritmo que analiza la entrada de los mensajes y las primeras palabras claves basadas en un conjunto de conceptos. • Se decide la influencia de conceptos establecidos en la ontología y la semántica y se crea un vector de pesos. • Se obtendrá una medida de similitud entre los mensajes de entrada y los resultados preliminares. • Desarrollo de una ontología de motor de búsqueda basado en el algoritmo WI-OUTOSEARCH.
Tesis	Con el fin de mejorar la precisión de las búsquedas mediante el análisis semántico el documento propone conceptos - peso de vectores y algoritmo de emparejamiento (CWVMA). El algoritmo en primer lugar, analiza los mensajes de entrada y las palabras clave preliminares basados en un conjunto de conceptos de resultados. Entonces decide la regla según la influencia de los conceptos establecidos en la ontología semántica y crea vector de peso, haciendo coincidir la regla. Por fin obtiene vector resultado como fundamento de la medida de similitud entre los mensajes de entrada y los resultados preliminares. Además, se diseña y se desarrolla un motor de búsqueda de ontología basada en el algoritmo anterior WI prototipo de sistema OntoSearch. El sistema puede buscar cerca de 4 millones de páginas web de Google Web Service. Muchos de los resultados de los experimentos de explicar el algoritmo puede mejorar la precisión de búsqueda de la Ontología
Tipo de investigación	Exploratoria
Metodología	
Tipo de metodología	
Técnicas	Algoritmo WI-OUTOSEARCH, vector de pesos, la captura de la semántica la realiza TUCUXI.
Resultados	-
Conclusiones	Este documento propone CWVMA de acuerdo a factores que la literatura presenta para medir la ontología semántica. Desarrollamos WI OntoSearch que es un motor de búsqueda basado en la ontología CWVMA y hace varios experimentos por él. Un gran número de experimentos con resultados muestran el algoritmo tiene un buen efecto en la mejora de la precisión de la búsqueda de la ontología.
Recomendaciones	El algoritmo con diferentes normas de congruencia, puede satisfacer diferentes necesidades. Otro trabajo interesante y atractivo es la agrupación ontologías por el algoritmo con las normas de la cartografía compleja.
Análisis del documento	
Resumen del documento	La búsqueda de información útil y la localización adecuada La ontología de la WWW o la Web Semántica es una tarea importante en el ámbito de la investigación Ontología. La diferencia entre la mayoría de los Información Ontología y la información común es que La ontología tiene una estructura semántica. Con el fin de mejorar la búsqueda precisión mediante el análisis semántico el documento propone conceptos - pesos

	<p>vectores algoritmo de emparejamiento (CWVMA). El en primer lugar, el algoritmo analiza los mensajes de entrada y preliminar Palabras clave de resultados basados en conjuntos de conceptos. Entonces decide corresponden a las reglas de acuerdo a la influencia de los conceptos establecidos en la Ontología y semántica crea vector de peso, haciendo coincidir regla. Por fin obtiene vector resultado como fundamento de la medida de similitud entre los mensajes de entrada y los resultados preliminares. En Además, este documento diseñado y desarrollado una ontología motor de búsqueda basado en el algoritmo anterior. El sistema puede buscar cerca de 4 millones de páginas web de Google Web Service. Una gran cantidad de resultados de experimentos explicar el algoritmo puede mejorar la precisión de la ontología de búsqueda.</p>
Palabras claves	Recuperación de información, ontologías, algoritmos.
Anexos	-
Glosas	-
Comentarios	-

Tabla 13. Ficha B 5 Recuperación de Información y Web Semántica.

2 ANEXO B

2.1 Artículo de Investigación

En esta sección se presenta el artículo de investigación “*Modelo Semántico de Expansión de Consulta para la Búsqueda Web - MSEC*”, el cual fue enviado a la “Revista Facultad de Ingeniería” de la Universidad de Antioquia.

Modelo Semántico de Expansión de Consultas para la Búsqueda Web (MSEC)

MIGUEL ANGEL NIÑO ZAMBRANO

Ingeniero de Sistemas, Magíster en Informática, Estudiante de Doctorado en Ingeniería de Sistemas y Computación
Profesor Titular, Departamento de Sistemas, Facultad de Ingeniería Electrónica y Telecomunicaciones
Miembro del Grupo de I+D en Tecnologías de la Información, Universidad del Cauca
manzamb@unicauca.edu.co

CARLOS ALBERTO COBOS LOZADA

Ingeniero de Sistemas, Magíster en Informática, Ph.D. (c) en Ingeniería de Sistemas y Computación
Profesor Titular, Departamento de Sistemas, Facultad de Ingeniería Electrónica y Telecomunicaciones
Director del Grupo de I+D en Tecnologías de la Información, Universidad del Cauca
ccobos@unicauca.edu.co

IVAN DARIO LOPEZ GOMEZ

Estudiante de Ingeniería de Sistemas
Programa de Ingeniería de Sistemas, Universidad del Cauca
Miembro del Grupo de I+D en Tecnologías de la Información, Universidad del Cauca
navis@unicauca.edu.co

CARLOS ADRIAN ANDRADE HOYOS

Estudiante de Ingeniería de Sistemas
Programa de Ingeniería de Sistemas, Universidad del Cauca
Miembro del Grupo de I+D en Tecnologías de la Información, Universidad del Cauca
caandrade@unicauca.edu.co

RESUMEN

Internet se ha convertido en el mayor repositorio de conocimiento humano y la cantidad de información almacenada crece cada día más. Esto repercute en el bajo nivel de precisión que reportan los sistemas de búsqueda Web respecto a los documentos que son recuperados al usuario, dificultando su tarea de búsqueda. Para enfrentar este problema es necesario proponer estrategias que se enfoquen en la recuperación personalizada de recursos. Este artículo describe un modelo semántico de expansión de consultas denominado MSEC, el cual está basado principalmente en la Similitud Semántica a partir de Ontologías de dominio y el uso del Perfil de Usuario con el fin de personalizar las búsquedas del usuario y así mejorar la precisión de las mismas. Un prototipo software ha sido desarrollado para evaluar el modelo propuesto y los resultados experimentales muestran una mejora del 3,5% respecto al enfoque tradicional de búsqueda.

3 ANEXO C

En este apartado se describen los principales aspectos que fueron tenidos en cuenta para la selección de la Ontología de dominio manejada por el presente proyecto.

3.1 Elección de la Ontología

Las Ontologías juegan un papel muy importante cuando se comparte conocimiento alrededor de un dominio particular, de esta forma, los profesionales de las distintas áreas del conocimiento, pueden establecer una semántica común respecto a un tema específico permitiendo una efectiva integración de la información [1].

Actualmente existen miles de Ontologías disponibles en internet, pero solo una pequeña porción de estas cumple con unas características adecuadas que posibilitan el aprovechamiento del componente semántico que las caracteriza. Para la elección de la Ontología que se utiliza en este proyecto, se llevó a cabo una amplia búsqueda en la Web, mediante la cual se analizaron diferentes Ontologías en el lenguaje OWL [2] debido a que este extiende RDFS para permitir la expresión de relaciones complejas entre diferentes clases RDFS, y brinda una mayor precisión en las restricciones de clases y propiedades específicas. En la primera etapa del proyecto no se había definido un dominio particular del conocimiento para encontrar una Ontología referente a este, de modo que a medida que se desarrollaba la etapa de búsqueda, se encontraron varias Ontologías del área de la medicina, cuya estructura describía ampliamente la semántica de dicho dominio y permitía hacer uso efectivo de sus componentes para la expansión de la consulta de usuario, por consiguiente se optó por establecer que el dominio específico a tratar sería la Oncología, rama de la medicina que estudia los tumores benignos y malignos, pero con especial atención a los malignos, esto es, al cáncer [3].

A continuación se presentan las Ontologías más destacadas en el dominio del Cáncer:

- **Ontología del Instituto Nacional del Cáncer.** La Ontología del Instituto Nacional del Cáncer [4] es una descripción lógica de dominio público basado en la terminología producida por el Instituto Nacional del Cáncer, distribuye como un componente del Centro de Bioinformática del NCI. Es profunda y compleja en comparación con vocabularios clínicos más amplios, la aplicación de relaciones ricas semánticas entre los nodos de su taxonomía. Las relaciones semánticas en la Ontología están destinadas a facilitar la investigación traslacional y apoyar la infraestructura de bioinformática del Instituto. Los temas descritos en la ontología son las enfermedades, medicamentos, productos químicos, diagnósticos, genes, tratamientos, anatomía, organismos, y proteínas.

Esta Ontología ha evolucionado a partir del Metatesauro del Instituto Nacional del Cáncer, el cual se basa en la Biblioteca Nacional de Medicina Unified Medical Language System (UMLS). El Metatesauro NCI ha estado en funcionamiento desde 1999. Una versión pública se encuentra disponible en <http://ncimeta.nci.nih.gov>. La traducción OWL actual del Tesoro NCI está disponible para descargar o ver en línea. Contiene algo más de 500.000 triplets.

- **OpenGALEN GRIAL y fuentes de OWL.** El modelo de referencia común OpenGALEN es una rica Ontología composicional del dominio médico que cubre anatomía, patología, enfermedades, síntomas, drogas y procedimientos. Esta ontología fue desarrollada en la Universidad de Manchester [5] y sus archivos originales están expresados en el lenguaje de descripción lógica GRAIL [6] junto con la documentación asociada, artículos académicos asociados y otras referencias están disponibles en www.opengalen.org.

Tomando como base estas dos Ontologías de dominio, se opta por utilizar la Ontología del Instituto Nacional del Cáncer de los Estados Unidos, puesto que su semántica es muy rica y además se encuentra unificada en un solo recurso OWL a diferencia de OpenGALEN que se encuentra distribuida en diferentes archivos y además de esto es una traducción a OWL-RDF a partir de GRAIL, lo cual no garantiza una correcta expresión de sus relaciones semánticas.

Con la decisión anterior, a continuación se anexa la licencia original de uso de esta Ontología.

3.1.1 Licencia Ontología NCI

NCI THESAURUS Terms of Use

NCI THESAURUSTM is a registered trademark of the National Cancer Institute.

Permission is hereby granted in perpetuity, without payment of license fees or royalties, to use, copy, or distribute the NCI THESAURUSTM ontology, NCI THESAURUSTM Users' Guide, and the NCI THESAURUSTM database for any commercial or noncommercial purpose, subject to the following terms and conditions:

- 1) To further the purpose of the NCI THESAURUS ontology and NCI THESAURUS database of providing a ontology and taxonomy specialized for modeling concepts relevant to cancer science, treatment and prevention, the only source of the NCI THESAURUS Users' Guide, the NCI THESAURUS ontology, its codes, and/or the NCI THESAURUS database shall be the National Cancer Institute.
- 2) Users may freely modify the content of the NCI THESAURUS ontology, database or documentation. However any such modified content may not be identified or represented as

being the NCI Thesaurus ontology, database or documentation, or part of the NCI Thesaurus ontology, database or documentation. Users are prohibited from using the NCI THESAURUS trademarks in conjunction with such modified content or work derived from the NCI THESAURUS, except to provide acknowledgement that the content was derived from the NCI THESAURUS.

4) NCI THESAURUS concept names, concept codes and other information from the NCI THESAURUS ontology and database may be used in tagging records in users' databases and electronic messages, without the need to include any NCI THESAURUS License. When the NCI THESAURUS concept identifier is used in tagging records in users' databases and electronic messages, users are encouraged, but not required, to include the corresponding NCI THESAURUS *Preferred Name* (from the *Preferred_Name* attribute) if the database or message provides a place for a text name representation of the code.

5) Users may make and distribute an unlimited number of copies of the unmodified NCI THESAURUS database. Each copy of the unmodified NCI THESAURUS database must include this NCI THESAURUS License, and must include the version number of the database. This NCI THESAURUS License must appear on every printed copy of the NCI THESAURUS database. Where the NCI THESAURUS database is distributed on a fixed storage medium (such as DVD-ROM or CD-ROM), a printed copy of this NCI THESAURUS License must be included on or with the storage medium, and a text file containing this information also must be stored on the storage medium in a file called "NCI_THESAURUS_license.txt". Where the NCI THESAURUS database is distributed via the Internet, this NCI THESAURUS License must be accessible on the same Internet page from which the NCI THESAURUS database is available for download.

6) Users may incorporate *in toto* the NCI THESAURUS database into another ontology or vocabulary database or software program. Every copy of the NCI THESAURUS database incorporated into or distributed in conjunction with another database or software program must include the following notice:

"This product includes the NCI THESAURUS™ database, or is derived from the NCI THESAURUS™ database, subject to a license from the National Cancer Institute. Your use of the NCI THESAURUS database and NCI THESAURUS ontology also is subject to this license, a copy of which is available at <ftp://ftp1.nci.nih.gov/pub/cacore/EVS/ThesaurusTermsOfUse.htm>. The current complete NCI THESAURUS database and Users' Guide are available for download at <ftp://ftp1.nci.nih.gov/pub/cacore/EVS/>. The NCI THESAURUS DATABASE IS PROVIDED "AS IS." ANY EXPRESS OR IMPLIED WARRANTIES ARE DISCLAIMED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. NCI THESAURUS™ is a registered trademark of the National Cancer Institute."

If the NCI THESAURUS database or software program containing the NCI THESAURUS database is distributed with a printed license, this statement must appear in the printed license. Where the NCI THESAURUS database or software program containing the NCI THESAURUS database is distributed on a fixed storage medium, a text file containing this

information also must be stored on the storage medium in a file called "NCI_THESAURUS_short_license.txt". Where the NCI THESAURUS or software program containing the NCI THESAURUS database is distributed via the Internet, this information must be accessible on the same Internet page from which the product is available for download.

7) Use and distribution of the NCI THESAURUS concept codes and NCI THESAURUS database in ways that are not specifically discussed herein shall be accompanied by the notice provided in Section 6 hereof. The guidelines for providing the notice that are contained in the last paragraph of Section 6 also shall apply. If a user has a question about whether a particular use of the NCI THESAURUS ontology and/or the NCI THESAURUS database is permissible, the user is invited to contact the National Cancer Institute by e-mail at **ncicb@pop.nci.nih.gov**.

8) The names "National Cancer Institute", "NCI" and "NCI Thesaurus" may not be used in a way that could be interpreted as an endorsement or a promotion of any product or service without prior written permission of the National Cancer Institute. To request written permission; please contact the National Cancer Institute by e-mail at **ncicb@pop.nci.nih.gov**.

DISCLAIMER: THE NATIONAL CANCER INSTITUTE DOES NOT ACCEPT LIABILITY FOR ANY OMISSIONS OR ERRORS IN THE NCI THESAURUS, NCI THESAURUS USERS' GUIDE OR THE NCI THESAURUS DATABASE. THE NCI THESAURUS, NCI THESAURUS USERS' GUIDE AND NCI THESAURUS DATABASE ARE PROVIDED "AS IS," WITHOUT WARRANTY OF ANY KIND.

ANY EXPRESS OR IMPLIED WARRANTIES ARE DISCLAIMED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL THE NATIONAL CANCER INSTITUTE OR CONTRIBUTORS TO THE NCI THESAURUS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF NCI THESAURUS USERS' GUIDE OR THE NCI THESAURUS DATABASE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

4 ANEXO D

4.1 Arquitectura de la Aplicación

Aquí se describen los patrones utilizados en la construcción del prototipo software y la arquitectura de la aplicación por medio del modelo “4+1 vistas” [7].

4.1.1 Vistas de la Arquitectura

Al desarrollar una aplicación es importante describir su arquitectura mediante una especificación de las decisiones estratégicas acerca de la estructura y la funcionalidad del sistema, las colaboraciones entre sus distintos elementos y su despliegue físico para cumplir unas responsabilidades bien definidas. El modelo “4+1” vistas describe la arquitectura del software usando cinco vistas concurrentes las cuales permiten a los diseñadores software organizar la descripción de sus decisiones de arquitectura en cuatro vistas y luego ilustrarlas con un conjunto reducido de casos de uso o escenarios, los cuales constituyen la quinta vista. La arquitectura evoluciona parcialmente a partir de estos escenarios [7].

4.1.1.1 Vista Conceptual

La vista conceptual es usada para definir los requerimientos funcionales y la visión que los usuarios del negocio tienen de la aplicación y describir el modelo de negocio que la arquitectura debe cubrir.

4.1.1.1.1 Diagrama de Casos de Uso

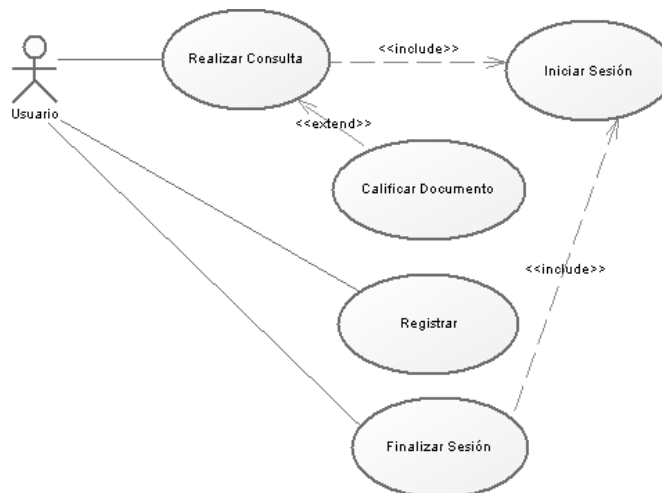


Figura 2. Diagrama de Casos de Uso.

4.1.1.1.2 Casos de Uso en formato extendido

CASO DE USO	Iniciar Sesión	
ACTOR	Usuario	
DESCRIPCIÓN	El usuario desea iniciar sesión para conectarse al sistema.	
PRECONDICIÓN	--	
CURSO NORMAL	ACTOR	SISTEMA
	1. Introduce su login y su clave. 2. Envía los datos al sistema.	3. Confirma la validez de los datos introducidos en el paso 1. 4. Inicia una sesión. 5. Muestra las funcionalidades que puede realizar el usuario.
CURSO ALTERNO	ACTOR	SISTEMA
		3a. Los datos introducidos en el paso 1 son inválidos: 4. No inicia sesión. 5. Envía mensaje a usuario notificando que el login y/o la clave son incorrectos.
POSTCONDICIÓN	Se creó una sesión para el usuario. El usuario está conectado al sistema y tiene acceso a todas sus funcionalidades.	

Tabla 14. Caso de Uso Iniciar Sesión.

CASO DE USO	Registrar	
ACTOR	Usuario	
DESCRIPCIÓN	El usuario desea registrarse en el sistema.	
PRECONDICIÓN		
CURSO NORMAL	ACTOR	SISTEMA
	1. Introduce sus datos personales. 2. Envía los datos al sistema. 4. Confirma la operación.	3. Solicita confirmación para registrar. 5. Guarda el registro del usuario 6. Envía mensaje notificando operación exitosa
CURSO ALTERNO	ACTOR	SISTEMA
	4a. Cancela la operación.	5. No se guarda el registro 6. Envía mensaje notificando operación fallida, pues existen tipos de datos que son erróneos.
POSTCONDICIÓN	El nuevo registro es guardado en el sistema.	

Tabla 15. Caso de Uso Registrar.

CASO DE USO	Realizar Consulta	
ACTOR	Usuario	
DESCRIPCIÓN	El usuario podrá realizar una consulta en la web, digitando la frase o frases que describan su pregunta. Con esto se pretende obtener una respuesta relevante a sus necesidades de información.	
PRECONDICIÓN	El usuario está conectado al sistema.	
CURSO NORMAL	ACTOR	SISTEMA
	1. introduce la(s) palabra(s) clave(s) para la búsqueda.	2. Muestra la lista de links con sus

		respectivos resúmenes que representan los documentos recuperados de acuerdo a la búsqueda.
CURSO ALTERNO	ACTOR	SISTEMA
	--	--
POSTCONDICIÓN	Se muestran los links con sus respectivos resúmenes que representan los documentos recuperados de acuerdo a la búsqueda.	

Tabla 16. Caso de Uso Realizar Consulta.

CASO DE USO	Calificar Documento	
ACTOR	Usuario	
DESCRIPCIÓN	El usuario desea ver el contenido del documento que ha seleccionado y posteriormente calificarlo.	
PRECONDICIÓN	El usuario está conectado al sistema. Debe haberse ejecutado el caso de uso Realizar Consulta	
CURSO NORMAL	ACTOR	SISTEMA
	1. Selecciona el link correspondiente al documento que pretende visualizar. 3. Selecciona la calificación que considera más apropiada sobre la relevancia del documento para su necesidad de información. 4. Envía la calificación al sistema.	2. Se muestra el contenido del documento seleccionado. 5. Guarda temporalmente la calificación del documento que ha sido enviada por el usuario 6. Retorna a la página que muestra los resultados de la búsqueda.
CURSO ALTERNO	ACTOR	SISTEMA
	--	5. No se guarda la calificación
POSTCONDICIÓN	Se guarda la calificación del documento seleccionado por el usuario.	

Tabla 17. Caso de Uso Calificar Documento.

CASO DE USO	Finalizar Sesión	
ACTOR	Usuario	
DESCRIPCIÓN	El usuario desea finalizar su sesión para desconectarse del sistema.	
PRECONDICIÓN	El usuario está conectado al sistema.	
CURSO NORMAL	ACTOR	SISTEMA
	1. El usuario selecciona la opción "Cerrar Sesión"	2. Obtiene las calificaciones de los documentos que el usuario ha seleccionado 3. Calcula los valores de Wru para cada concepto identificados en la consulta expandida 4. Guarda los conceptos con su respectivo Wru en el Perfil de Usuario 5. Des autentica al usuario logeado
CURSO ALTERNO	ACTOR	SISTEMA
	--	--
POSTCONDICIÓN	El usuario se desconecta del sistema y se guardan los conceptos de la consulta expandida con su respectivo Wru en el Perfil de Usuario	

Tabla 18. Caso de Uso Finalizar Sesión.

4.1.1.2 Vista Lógica

Muestra los componentes principales de diseño y sus relaciones de forma independiente de los detalles técnicos y de cómo la funcionalidad será implementada en la plataforma de ejecución (Figura 3).

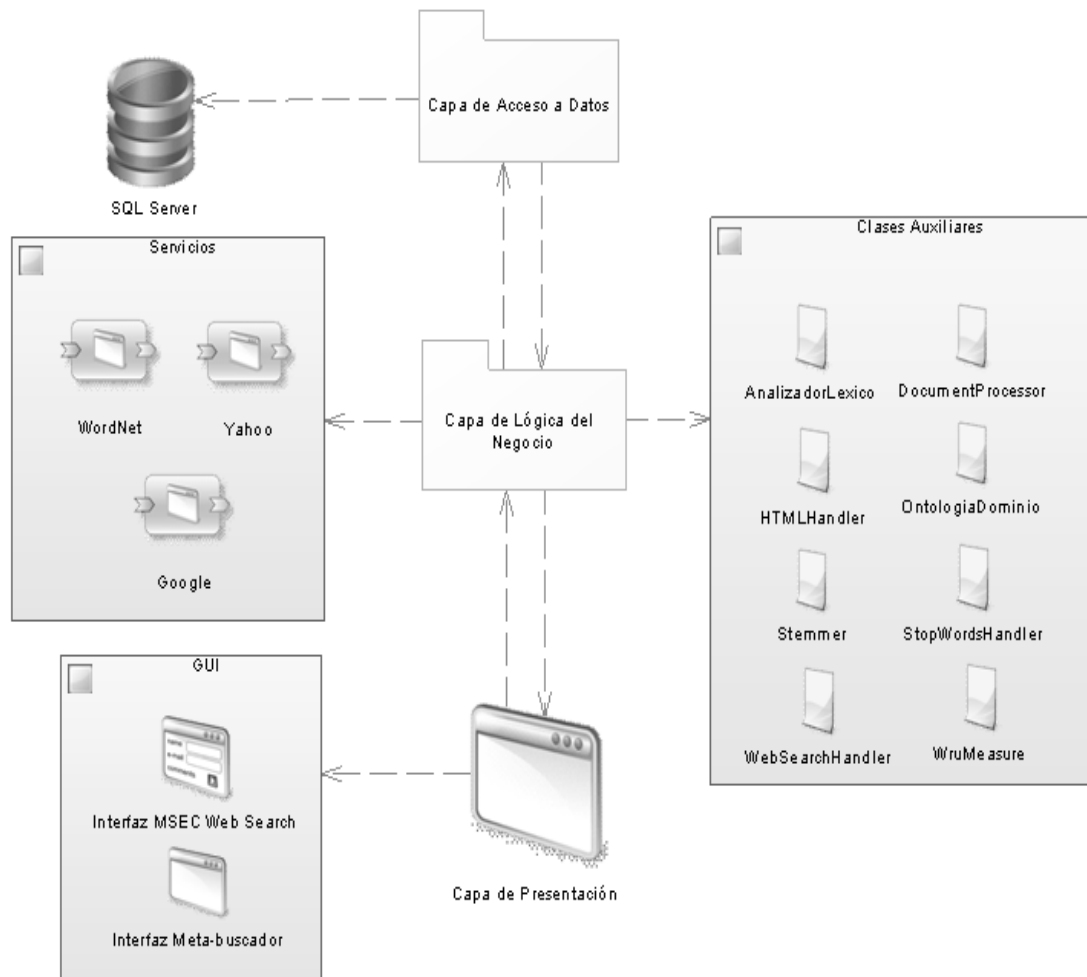


Figura 3. Vista lógica por capas de la aplicación.

4.1.1.3 Vista Física

Esta vista ilustra la distribución del procesamiento entre los distintos equipos que conforman la solución, incluyendo los servicios y procesos de base. Los elementos definidos en la vista lógica se "mapean" a componentes de software (servicios, procesos, etc.) o de hardware que definen más precisamente como se ejecutará la solución. La vista física del sistema se puede ver en la Figura 4 a través del diagrama de despliegue.

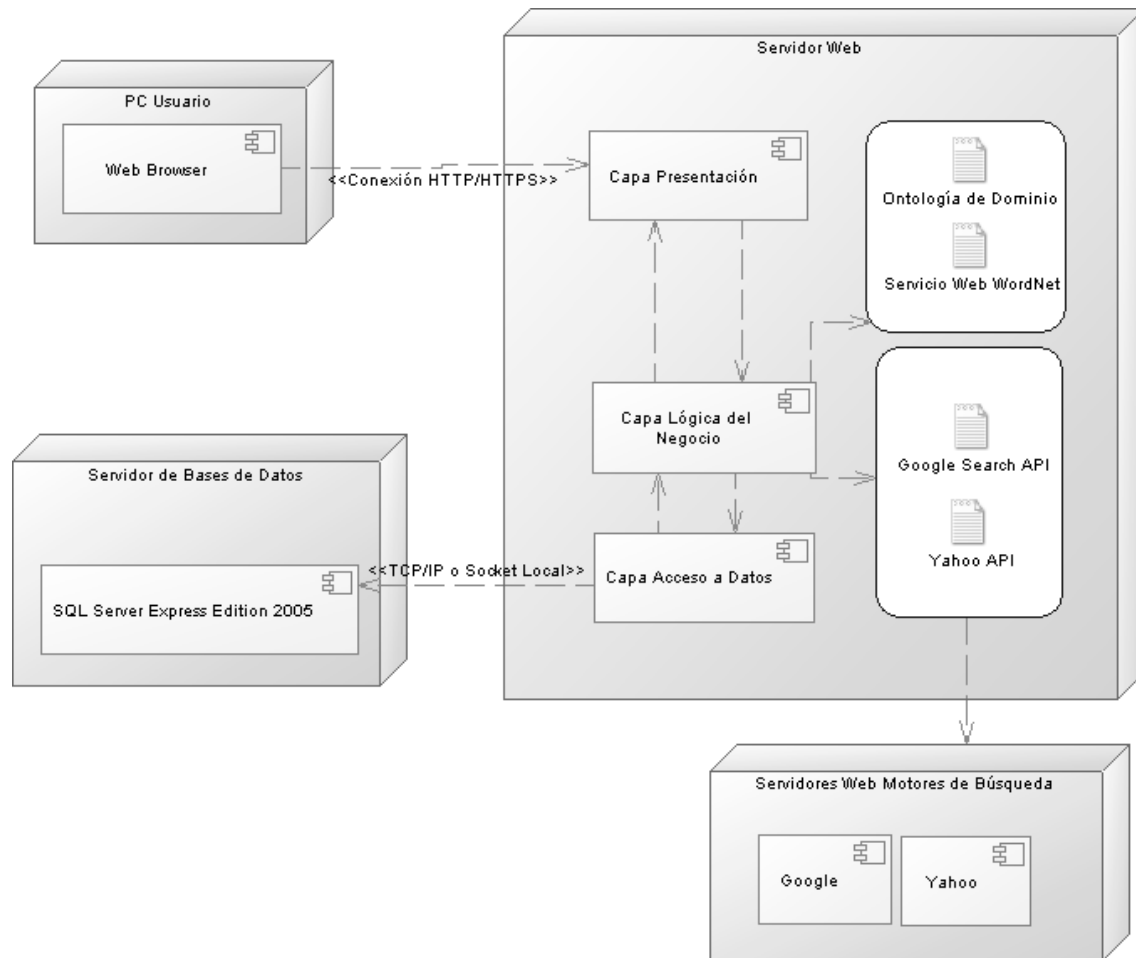


Figura 4. Diagrama de Despliegue.

4.1.1.4 Vista de Implementación

Finalmente la vista de implementación describe cómo se implementan los componentes físicos mostrados en vista de distribución agrupándolos en subsistemas organizados en capas y jerarquías, ilustra, además las dependencias entre éstos. Básicamente, se describe el mapeo desde los paquetes y clases del modelo de diseño a subsistemas y componentes físicos. La vista de implementación puede verse en la Figura 5.

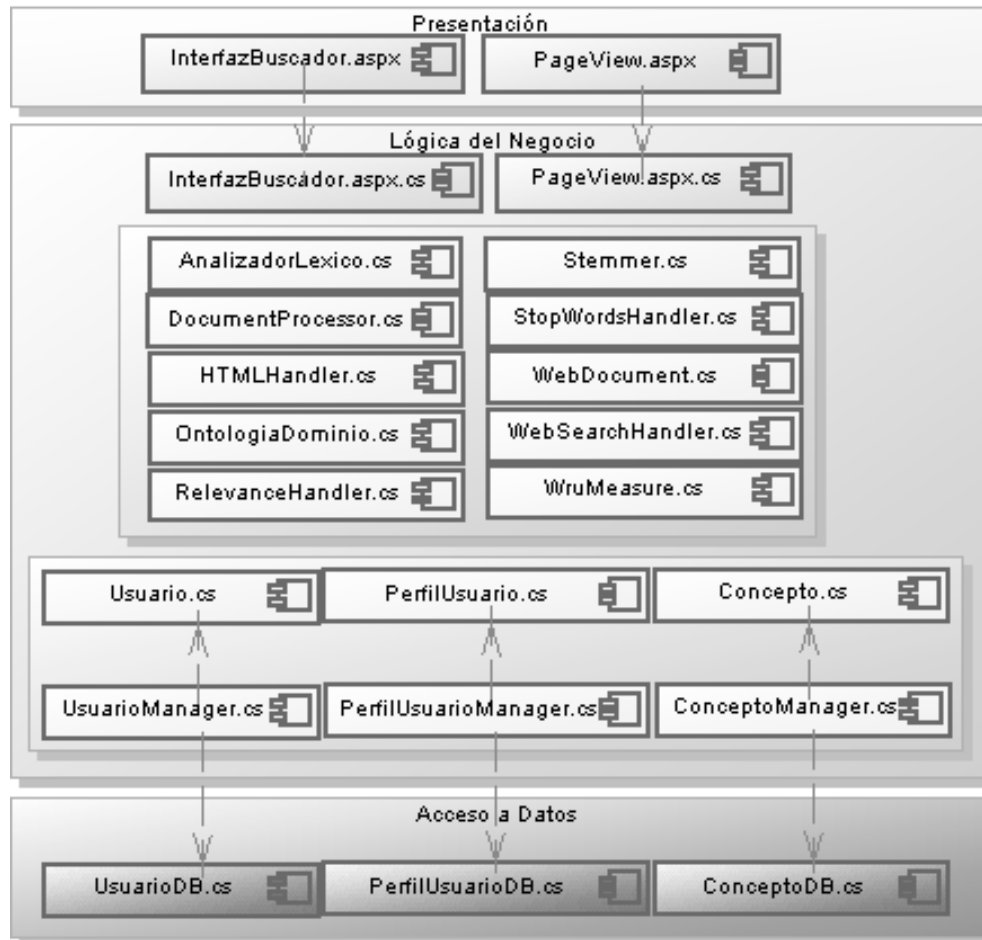


Figura 5. Vista de Implementación.

4.1.2 Patrones utilizados

Los patrones de diseño son el esqueleto de las soluciones a problemas comunes en el desarrollo de software [8], en otras palabras, brindan una solución ya probada y documentada a problemas de desarrollo de software que están sujetos a contextos similares.

- **Patrón Fachada.** Simplifica la interface entre dos sistemas o componentes software ocultando un sistema complejo detrás de una clase que hace las veces de pantalla o fachada. Su objetivo principal es ocultar la complejidad de un sistema, el conjunto de clases o componentes que lo conforman, de modo que solo se ofrezca un punto de entrada al sistema. Una de las ventajas de usar el patrón fachada para comunicar las dos partes o componentes, es la de aislar los posibles cambios que se puedan producir en alguna de las partes.
- **Patrón Code Behind.** El patrón utilizado en la plataforma .NET es el patrón Code-Behind el cual separa la interfaz grafica de la lógica, dividiendo en compilación el

tiempo de ejecución y compilación previa. Este patrón permite dividir la interfaz del usuario con la definición de los objetos, además tiene una clase que desarrolla la lógica de negocio de cada control o formulario.

5 ANEXO E

5.1 Prueba de Rendimiento

Para asegurar y garantizar el correcto funcionamiento del sistema se realizaron ciertas pruebas las cuales se seleccionaron para determinar si el sistema cumple con su función principal, brindar un servicio a los usuarios de manera eficiente y sin fallas o retrasos. Es importante tener en cuenta que el rendimiento del sistema esta estrictamente ligado con el hardware de la computadora que realiza la función de servidor, ya que esta determina el desempeño del mismo.

Con las pruebas a realizar también se pretende comprobar si existe un error o algún retraso que pueda afectar el funcionamiento del sistema, ya que se requiere que sea rápido, confiable y estable. Además todo debe ser transparente para el usuario, en caso de que exista un error grave en el sistema se debe informar al usuario por medio una interfaz amigable y entendible u otro medio la situación de dicha excepción.

5.1.1 Software y Hardware utilizados para las pruebas

Las pruebas fueron realizadas utilizando el software y hardware que se describe a continuación:

- **Procesador:** Intel(R) Core (TM) 2 Duo CPU E4500 @ 2.20 Ghz 2.19 Ghz.
- **Cache del procesador:** 16MB de caché de nivel 2.
- **Ram:** 4096 MB (4 dimm).
- **Hard disk:** 250GB (5400 RPM).
- **Sistema operativo:** Windows 7 32 bits.

Es de importancia mencionar que las características anteriores se utilizaron para desarrollar el software, y por ende utilizado para realizar pruebas de rendimiento del sistema. Para la realización de las pruebas se llevaron a cabo algunas tareas y se inspeccionó que el sistema funcionara y respondiera correctamente a todas las tareas. Este proceso en este sistema se realizo llevando a cabo 5 peticiones por usuario, las cuales están dadas en el siguiente orden.

- **Registro de Usuario:** En donde el usuario registra su login y password.
- **Inicio de sesión:** Logeo en el sistema, que consta de ingresar el login y password y desplegar la pagina de búsqueda de consultas.
- **Búsqueda:** Se ingresa la consulta, desplegando los resultados.

- **Cargue de página seleccionada:** Se selecciona la página la cual es desplegada y lista a calificar por el usuario.
- **Enviar Calificación:** Se envían al servidor las calificaciones.

Con base en lo anterior se construye un informe generado por 6 usuarios los cuales se conectan simultáneamente y ejecutan cada petición. Este informe está conformado por la Tabla 19 en donde las filas representan las 5 peticiones o actividades que se realizaron y las columnas indican ciertas medidas dentro de las pruebas las cuales se mencionan a continuación.

- **URL.** Es la actividad que se desempeña, el request o la petición.
- **# Muestras.** Es la cantidad de veces que se realizó la actividad (Una vez por cada usuario).
- **Media.** El promedio o media aritmética del tiempo en milisegundos.
- **Mediana.** Mediana del tiempo en milisegundos.
- **Min.** Tiempo mínimo de todas las peticiones de este tipo.
- **Max.** Tiempo máximo de todas las peticiones de ese tipo.
- **Porcentaje de error.** En el cual se muestra el porcentaje de las peticiones fallidas.
- **Rendimiento.** Esta medido en petición/segundo.
- **KB/Sec.** Medida de velocidad de Kilobytes/sec.

URL	# Muestras	Media	Mediana	Min	Max	% Error	Rendimiento	Kb/Sec
Registro de Usuario	6	1	0,01	0,02	0.05	0.00%	4.2/sec	1.4
Inicio de sesión	6	1	0.03	0.01	0.02	0.00%	5.8/sec	1.8
Búsqueda	6	50	15	5	11	0.00%	9.6/sec	5.93
Cargue de página seleccionada	6	60	13	6	11	0.00%	10.5%	2.4
Enviar Calificación	6	100	19	8	14	0.00%	15.9/sec	6.48
Total	30	212	46.07	19.03	36.07	0.00%	46/sec	18.01

Tabla 19. Prueba de rendimiento del sistema para 6 usuarios.

Se puede apreciar que la media fue de 212 ms, esto quiere decir que el sistema en promedio tarda en responder 0.2 segundos, lo cual es un tiempo bastante bueno tomando en cuenta que son 6 usuarios conectados al mismo tiempo lo que clasifica al sistema como rápido. Además también se muestra que el porcentaje de error en todas las peticiones es de 0.00%, esto indica que no fue desplegada ninguna página de error, ya que todas las peticiones fueron respondidas de manera adecuada. También es de importancia resaltar que cada petición realizada por el usuario considera en los tiempos de respuesta el acceso a la base de datos y a la ontología de dominio por lo que se concluye que el sistema se tarda muy poco tiempo en realizar el acceso a la base de datos y a la ontología para recuperar la información necesaria.

5.2 Prueba de Usabilidad

La usabilidad del sistema se evaluó mediante una prueba realizada a 6 estudiantes de noveno semestre del programa de fisioterapia de la Universidad del Cauca. La prueba está basada en una serie de preguntas para las cuales respondieron cada uno de los usuarios, según su criterio, cómo les parecía el sistema de acuerdo a su uso, visibilidad y demás factores de usabilidad, también teniendo en cuenta las observaciones y/o sugerencias que querían hacer al sistema.

Los factores que se tuvieron en cuenta, se muestran a continuación como en el documento entregado a los estudiantes de dicho colegio. Además, se consigna el número de personas que coincidieron en cada aspecto del grado de satisfacción (excelente, bueno, neutro, regular, deficiente) sobre el sistema.

A continuación en la Tabla 20 se presenta la prueba de usabilidad que fue aplicada a varios usuarios para determinar el comportamiento del sistema respecto a la interacción de los usuarios con el mismo.

Visibilidad del estado del sistema	Excelente	Bueno	Neutro	Regular	Deficiente
1.1. El sitio muestra claramente dónde se encuentra el usuario	3	2	1		
1.2. Los enlaces posibles de explorar están claramente señalados	2	4			

Relación entre sistema y mundo real	Excelente	Bueno	Neutro	Regular	Deficiente
2.1 El lenguaje es claro	4	1	1		
2.2. Los conceptos utilizados son entendibles	3	3			
2.3. Las palabras son de significado conocido	5		1		
2.4. Los iconos generan significado	2	2	2		

Consistencia y estándares	Excelente	Bueno	Neutro	Regular	Deficiente
3.1 Existe coherencia entre el nombre de un enlace y el sitio al que apunta	4	2			
3.2 Todos los enlaces tienen contenido	5	1			

Reconocer en lugar de recordar	Excelente	Bueno	Neutro	Regular	Deficiente
4.1 Los iconos son fácilmente reconocibles	3	2		1	

4.2. Los enlaces pueden identificarse claramente	3	1	2		
4.3. Es posible reconocer dónde se encuentra el usuario	2	3		1	

Recuperación de Información	Excelente	Bueno	Neutro	Regular	Deficiente
5.1 Encontró todos los conceptos (sobre medicina) buscados en las páginas retornadas.	3	1	2		
5.2. El buscador es fácil de usar.	5	1			
5.3. Los resultados arrojados cumplieron las expectativas de su búsqueda.	4	2			

Ayuda y documentación	Excelente	Bueno	Neutro	Regular	Deficiente
6.1 Existe algún tipo de ayuda o indicación en el sitio	6				
6.2. Cuando existe ayuda, ésta es específica	2	1		3	
6.3. La ayuda es asequible	5	1			

¿Cómo califica globalmente el sitio Web analizado?	Excelente	Bueno	Neutro	Regular	Deficiente
1. Visibilidad del estado del sistema	4	2			
2. Relación entre sistema y mundo real	4	1	1		
3. Control del usuario y libertad	2	2	1	1	
4. Prevención de errores	2	1	1		2
5. Flexibilidad y eficiencia de uso	3	3			
6. Estética y diseño	4	2			
7. Ayuda y documentación	4	1	1		

Tabla 20. Evaluación de usabilidad de la aplicación.

Al realizar la ponderación de respuestas se encontraron buenos resultados, pues la mayoría de personas estuvo de acuerdo en la calificación del sitio entre bueno y excelente, teniendo en cuenta aspectos como Visibilidad del estado y el sistema, recuperación de información y calificación global del sitio. A continuación se muestra en porcentajes los resultados obtenidos (Tabla 21).

Pregunta	Excelente	Bueno	Neutro	Regular	Deficiente
Pregunta 1.1	50	33,3	16,6	0	0
Pregunta 1.2	33,3	66,6	0	0	0
Pregunta 2.1	66,6	16,6	16,6	0	0
Pregunta 2.2	50	50	0	0	0

Pregunta 2.3	83,3	0	16,6	0	0
Pregunta 2.4	33,3	33,3	33,3	0	0
Pregunta 3.1	66,6	33,3	0	0	0
Pregunta 3.2	83,3	16,6	0	0	0
Pregunta 4.1	50	33,3	0	16,6	0
Pregunta 4.2	50	16,6	33,3	0	0
Pregunta 4.3	33,3	50	0	16,6	0
Pregunta 5.1	50	16,6	33,3	0	0
Pregunta 5.2	83,3	16,6	0	0	0
Pregunta 5.3	66,6	33,3	0	0	0
Pregunta 6.1	100	0	0	0	0
Pregunta 6.2	33,3	16,6	0	50	0
Pregunta 6.3	83,3	16,6	0	0	0
Pregunta I	66,6	33,3	0	0	0
Pregunta II	66,6	16,6	16,6	0	0
Pregunta III	33,3	33,3	16,6	16,6	0
Pregunta IV	33,3	16,6	16,6	0	33,3
Pregunta V	50	50	0	0	0
Pregunta VI	66,6	33,3	0	0	0
Pregunta VII	66,6	16,6	16,6	0	0

Tabla 21. Porcentaje de resultados para cada pregunta de usabilidad.

Como se observa, los principales porcentajes según la Tabla 21, se encuentran en las calificaciones de bueno y excelente. Esto permite observar una buena apreciación en cuanto a la usabilidad que presenta la aplicación y relevancia, en general, de los resultados obtenidos de algunas consultas.

6 ANEXO F

6.1 Validación del Prototipo

En esta sección se describe la fase de validación del prototipo aplicando medidas como la curva Precision-Recall, Índice MAP y estadísticas Kappa.

6.1.1 Curva Precisión-Recall

En la primera etapa de las pruebas de relevancia se llevan a cabo estas medidas sobre colecciones cerradas de documentos, donde la relevancia de cada documento de acuerdo a una consulta específica, ha sido previamente evaluada por uno o más expertos en el dominio. Se opta por utilizar la colección *MED Collection* [9] compuesta por documentos cuya temática es la Medicina, debido a que en el prototipo software desarrollado en este proyecto se recurre a una Ontología referente al cáncer en el dominio de las ciencias de la salud.

En la realización de las pruebas se seleccionaron 15 consultas de la colección de documentos de prueba la cual está compuesta por 30 consultas, cada una de estas se envía al meta-buscador MSEC Web Search para realizar la expansión de las mismas y posteriormente permitir la búsqueda de los documentos relevantes sobre la colección de documentos¹.

A continuación se muestran las pruebas de relevancia con una colección de documentos para tres consultas a modo de ejemplo. Se calcula la precisión, el recuerdo y la precisión promedio teniendo en cuenta cada uno de los cinco formatos de texto aplicados a cada consulta.

- **Identificador de Consulta²: 22**

Consulta sin Expansión: mycoplasma (infection or presence) in embryo, fetus, newborn infant or animal, or in pregnancy, gynecologic diseases, or as related to chromosomes or chromosome abnormality, or microanatomy

Nº de documentos recuperados	Nº Documentos Relevantes para el usuario	Recall	Precisión
1	1	0,04	1
3	2	0,08	0,66666667
6	3	0,12	0,5
10	4	0,16	0,4

¹ En esta primera etapa de pruebas de relevancia se ha indexado MED Collection a través de Lucene .NET y la búsqueda se realiza de forma local, es decir sin el uso de los buscadores Web Google y Yahoo.

² Este identificador es propio de la consulta y viene por defecto en la colección.

13	5	0,2	0,38461538
16	6	0,24	0,375
18	7	0,28	0,38888889
21	8	0,32	0,38095238

Tabla 22. Cálculo de Precisión y Recall Consulta 22 - Sin expansión

Precisión Promedio: 0,512015415

Consulta Expandida con operadores OR AND (sin PU): mycoplasma OR infection OR presence OR diseases OR disease OR chromosomes OR chromosome OR abnormaliti OR (Embryo) OR (Fetus AND (Uterus OR Nucleus)) OR (Newborn) OR (Infant) OR (Animal) OR (Pregnancy) OR (Gynecologic) OR (Related) OR (Chromosome) OR (Microanatomy) OR Nuclear Structure OR Subcellular Structure OR Cell Structure.

Tipos de conceptos identificados:

CS → Embryo, Newborn, Infant, Animal, Pregnancy, Gynecologic, Related, Chromosome, Microanatomy
(CS → Fetus, CSR → Uterus, Nucleus)
CSE → Nuclear Structure, Subcellular Structure, Cell Structure (conceptos entre "Chromosome" y "Microanatomy")
T → mycoplasma, infection, presence, abnormality
(T → diseases, SIN → disease), (T → chromosomes, SIN → chromosome),

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
1	1	0,04	1
4	2	0,08	0,5
8	3	0,12	0,375
11	4	0,16	0,36363636
15	5	0,2	0,33333333
17	6	0,24	0,35294118
19	7	0,28	0,36842105
23	8	0,32	0,34782609

Tabla 23. Cálculo de Precisión y Recall Consulta 22 - Operadores OR AND (sin PU)

Precisión Promedio: 0,455144752

Consulta Expandida con operadores OR AND (con PU): mycoplasma OR infection OR presence OR diseases OR disease OR chromosomes OR chromosome OR abnormaliti OR (Embryo) OR (Fetus AND (Uterus OR Nucleus)) OR (Newborn) OR (Infant) OR (Animal) OR (Pregnancy) OR (Gynecologic) OR (Related) OR (Chromosome) OR (Microanatomy) OR Subcellular Structure.

Tipos de conceptos identificados:

CS → Embryo, Newborn, Infant, Animal, Pregnancy, Gynecologic, Related, Chromosome, Microanatomy
(CS → Fetus, CSR → Uterus, Nucleus)

CSE → Subcellular Structure con wru mayor (concepto entre “Chromosome” y “Microanatomy”)
T → mycoplasma, infection, presence, abnormaliti
(T → diseases, SIN → disease), (T → chromosomes, SIN → chromosome),

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
1	1	0,04	1
2	2	0,08	1
5	3	0,12	0,6
7	4	0,16	0,57142857
9	5	0,2	0,55555556
11	6	0,24	0,54545455
13	7	0,28	0,53846154
15	8	0,32	0,53333333
18	9	0,36	0,5

Tabla 24. Cálculo de Precisión y Recall Consulta 22 - Operadores OR AND (con PU)

Precisión Promedio: 0,649359283

Consulta Expandida con operadores OR (sin PU): mycoplasma OR infection OR presence OR diseases OR disease OR chromosomes OR chromosome OR abnormaliti OR (Embryo) OR (Fetus OR (Uterus OR Nucleus)) OR (Newborn) OR (Infant) OR (Animal) OR (Pregnancy) OR (Gynecologic) OR (Related) OR (Chromosome) OR (Microanatomy) OR Nuclear Structure OR Subcellular Structure OR Cell Structure.

Tipos de conceptos identificados:

CS → Embryo, Newborn, Infant, Animal, Pregnancy, Gynecologic, Related, Chromosome, Microanatomy
(CS → Fetus, CSR → Uterus, Nucleus)
CSE → Nuclear Structure, Subcellular Structure, Cell Structure (conceptos entre “Chromosome” y “Microanatomy”)
T → mycoplasma, infection, presence, abnormaliti
(T → diseases, SIN → disease), (T → chromosomes, SIN → chromosome),

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
2	1	0,04	0,5
6	2	0,08	0,33333333
10	3	0,12	0,3
14	4	0,16	0,28571429
17	5	0,2	0,29411765

Tabla 25. Cálculo de Precisión y Recall Consulta 22 - Operador OR (sin PU)

Precisión Promedio: 0,342633053

Consulta Expandida con operadores OR (con PU): mycoplasma OR infection OR presence OR diseases OR disease OR chromosomes OR chromosome OR

abnormaliti OR (Embryo) OR (Fetus OR (Uterus OR Nucleus)) OR (Newborn) OR (Infant) OR (Animal) OR (Pregnancy) OR (Gynecologic) OR (Related) OR (Chromosome) OR (Microanatomy) OR Subcellular Structure.

Tipos de conceptos identificados:

CS → Embryo, Newborn, Infant, Animal, Pregnancy, Gynecologic, Related, Chromosome, Microanatomy
(CS → Fetus, CSR → Uterus, Nucleus)
CSE → Subcellular Structure con wru mayor (concepto entre “Chromosome” y “Microanatomy”)
T → mycoplasma, infection, presence, abnormaliti
(T → diseases, SIN → disease), (T → chromosomes, SIN → chromosome),

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
2	1	0,04	0,5
5	2	0,08	0,4
8	3	0,12	0,375
12	4	0,16	0,33333333
15	5	0,2	0,33333333
19	6	0,24	0,31578947

Tabla 26. Cálculo de Precisión y Recall Consulta 22 - Operador OR (con PU)

Precisión Promedio: 0,37624269

RECALL	PRECISION				
	Expansión con operadores (OR AND) y PU	Expansión con operadores (OR AND)	Consulta sin Expansión	Expansión con operadores (OR) y PU	Expansión con operadores (OR)
0.04	1	1	1	0.5	0.5
0.08	1	0.5	0.666666667	0.4	0.333333333
0.12	0.6	0.375	0.5	0.375	0.3
0.16	0.571428571	0.363636364	0.4	0.333333333	0.285714286
0.2	0.555555556	0.333333333	0.384615385	0.333333333	0.294117647
0.24	0.545454545	0.352941176	0.375	0.315789474	
0.28	0.538461538	0.368421053	0.388888889		
0.32	0.533333333	0.347826087	0.380952381		
0.36	0.5				

Tabla 27. Consulta 22 - Comparación curva Precisión–Recall de la consulta sin expansión y la expansión de consulta con operadores or y operadores or-and evaluadas con y sin PU.

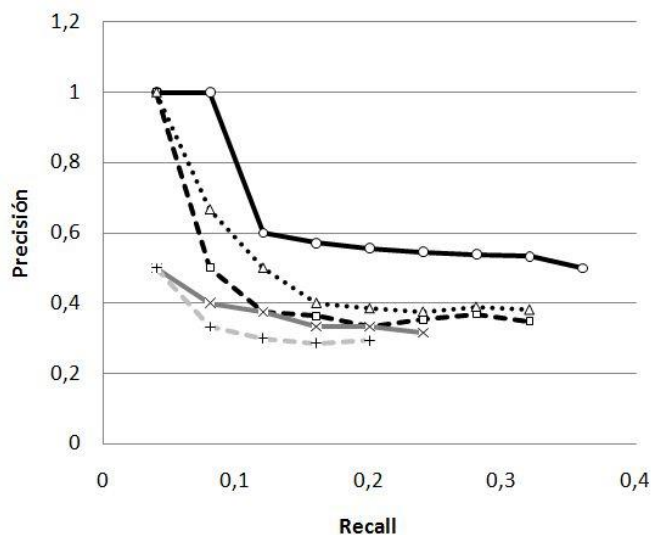


Figura 6. Curvas Precisión-Recuerdo para la primera consulta.

En la gráfica anterior mostrada en la Figura 6 se observa que la curva con mejor comportamiento para valores de precisión es la correspondiente a la consulta expandida mediante la combinación de operadores *OR-AND* haciendo uso del PU, se muestra una diferencia significativa respecto a las demás curvas, además de esto, presenta valores de 1 hasta alcanzar un 8% en el nivel de recall, lo cual indica que se han obtenido documentos relevantes justamente en las primeras posiciones del ranking que es donde usualmente se el usuario centra su atención [10]. A medida que se recuperan documentos, esta curva decrece pero aún así, se mantiene por encima de las demás y sus valores de precisión se ubican en el rango de 0.6 – 0.5 alcanzando finalmente un nivel máximo de recall del 36%.

La segunda curva con mejor comportamiento pertenece a la consulta sin expansión, la cual se ha establecido como la curva de referencia para realizar las comparaciones. Esta curva presenta un buen comportamiento al inicio de la recuperación, sin embargo los valores de precisión decaen rápidamente y a medida que se recuperan más documentos relevantes, estos valores no superan el 40%, por consiguiente, esta curva se ve superada por más de un punto de precisión por la consulta expandida con operadores *OR-AND* con PU.

Las demás curvas presentan un comportamiento no tan bueno comparado con el de las anteriores dos, si bien la curva correspondiente a la consulta expandida con operadores *OR-AND* sin PU alcanza el valor inicial 1 en la precisión, este decae muy rápidamente hasta alcanzar el valor 0.5 y a lo largo de la recuperación de documentos siempre se encuentra por debajo de la curva de referencia al igual que la curvas de la consulta expandida mediante el operador *OR* y la curva de la consulta expandida con este mismo operador con el PU, estas curvas no presentan un buen comportamiento al inicio de la recuperación debido a que sólo alcanzan un 50% de precisión, es decir

los documentos relevantes recuperados no se ubican en las primeras posiciones del ranking.

- **Identificador de Consulta:** 26

Consulta sin Expansión: method for experimental production of and known cause of hydrocephalus in animal and human

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
1	1	0.035714286	1
3	2	0.071428571	0.66666667
6	3	0.107142857	0.5
8	4	0.142857143	0.5
9	5	0.178571429	0.55555556
11	6	0.214285714	0.54545455
14	7	0.25	0.5
17	8	0.285714286	0.47058824
20	9	0.321428571	0.45
23	10	0.357142857	0.43478261

Tabla 28. Cálculo de Precisión y Recall Consulta 26 - Sin expansión

Precisión Promedio: 0.562304761

Consulta Expandida con operadores OR AND (sin PU): method OR production OR known OR cause OR (Experimental) OR (Hydrocephalus AND (Nervous System)) OR (Animal) OR (Human) OR Primate OR Mammal OR Vertebrate.

Tipos de conceptos identificados:

CS → Experimental, Animal y Human
(CS → Hydrocephalus, CSR → Nervous System)
CSE → Primate, Mammal, Vertebrate (conceptos entre "Human" y "Animal")
T → method, production, known y cause

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
1	1	0.03571429	1
4	2	0.07142857	0.5
7	3	0.10714286	0.428571429
9	4	0.14285714	0.444444444
11	5	0.17857143	0.454545455
14	6	0.21428571	0.428571429
18	7	0.25	0.388888889
22	8	0.28571429	0.363636364
26	9	0.32142857	0.346153846

Tabla 29. Cálculo de Precisión y Recall Consulta 26 - Operadores OR AND (sin PU)

Precisión Promedio: 0.483867984

Consulta Expandida con operadores OR AND (con PU): method OR production OR known OR cause OR (Experimental) OR (Hydrocephalus AND (Nervous System)) OR (Animal) OR (Human) OR Vertebrate.

Tipos de conceptos identificados:

CS → Experimental, Animal y Human
(CS → Hydrocephalus, CSR → Nervous System)
CSE → Vertebrate con wru mayor (concepto entre "Human" y "Animal")
T→ method, production, known y cause

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
1	1	0.035714286	1
3	2	0.071428571	0.66666667
5	3	0.107142857	0.6
7	4	0.142857143	0.57142857
8	5	0.178571429	0.625
10	6	0.214285714	0.6
13	7	0.25	0.53846154
16	8	0.285714286	0.5
19	9	0.321428571	0.47368421
21	10	0.357142857	0.47619048
24	11	0.392857143	0.45833333
27	12	0.428571429	0.44444444

Tabla 30. Cálculo de Precisión y Recall Consulta 26 - Operadores OR AND (con PU)

Precisión Promedio: 0.579517437

Consulta Expandida con operadores OR (sin PU): method OR production OR known OR cause OR (Experimental) OR (Hydrocephalus OR (Nervous System)) OR (Animal) OR (Human) OR Primate OR Mammal OR Vertebrate.

Tipos de conceptos identificados:

CS → Experimental, Animal y Human
(CS → Hydrocephalus, CSR → Nervous System)
CSE → Primate, Mammal, Vertebrate (conceptos entre "Human" y "Animal")
T→ method, production, known y cause

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
1	1	0.035714286	1
6	2	0.071428571	0.33333333
10	3	0.107142857	0.3
13	4	0.142857143	0.30769231
16	5	0.178571429	0.3125
18	6	0.214285714	0.33333333

22	7	0.25	0.31818182
----	---	------	------------

Tabla 31. Cálculo de Precisión y Recall Consulta 26 - Operador OR (sin PU)

Precisión Promedio: 0.415005828

Consulta Expandida con operadores OR (con PU): method OR production OR known OR cause OR (Experimental) OR (Hydrocephalus OR (Nervous System)) OR (Animal) OR (Human) OR Vertebrate.

Tipos de conceptos identificados:

CS → Experimental, Animal y Human
(CS → Hydrocephalus, CSR → Nervous System)
CSE → Vertebrate con wru mayor (concepto entre "Human" y "Animal")
T → method, production, known y cause

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
1	1	0.035714286	1
5	2	0.071428571	0.4
9	3	0.107142857	0.33333333
12	4	0.142857143	0.33333333
14	5	0.178571429	0.35714286
17	6	0.214285714	0.35294118
20	7	0.25	0.35
24	8	0.285714286	0.33333333

Tabla 32. Cálculo de Precisión y Recall Consulta 26 - Operador OR (con PU)

Precisión Promedio: 0.432510504

RECALL	PRECISION				
	Expansión con operadores (OR AND) y PU	Expansión con operadores (OR AND)	Consulta sin Expansión	Expansión con operadores (OR) y PU	Expansión con operadores (OR)
0.03571429	1	1	1	1	1
0.07142857	0.666666667	0.5	0.666666667	0.4	0.33333333
0.10714286	0.6	0.428571429	0.5	0.33333333	0.3
0.14285714	0.571428571	0.444444444	0.5	0.33333333	0.307692308
0.17857143	0.625	0.454545455	0.555555556	0.357142857	0.3125
0.21428571	0.6	0.428571429	0.545454545	0.352941176	0.33333333
0.25	0.538461538	0.388888889	0.5	0.35	0.318181818
0.28571429	0.5	0.363636364	0.470588235	0.33333333	
0.32142857	0.473684211	0.346153846	0.45		
0.35714286	0.476190476		0.434782609		
0.39285714	0.458333333				
0.42857143	0.444444444				

Tabla 33. Consulta 26 - Comparación curva Precision-Recall de la consulta sin expansión y la expansión de consulta con operadores OR y operadores OR-AND evaluadas con y sin PU.

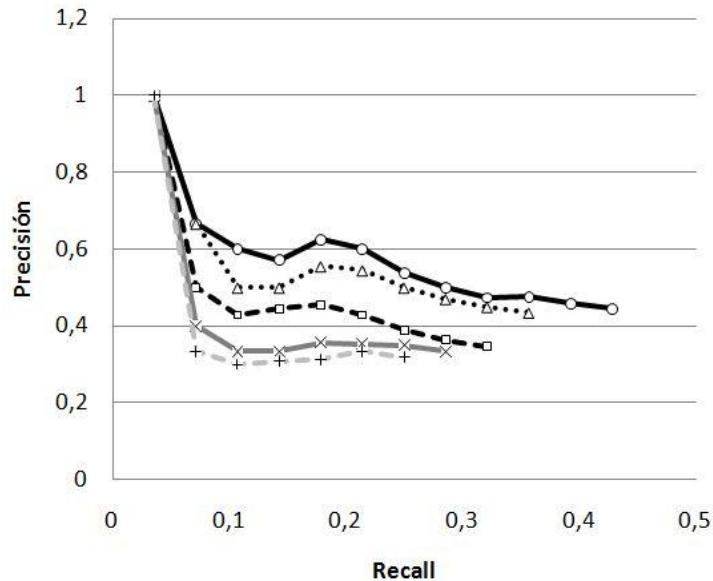


Figura 7. Curvas Precisión-Recuerdo para la segunda consulta.

En la gráfica anterior mostrada en la Figura 7 se puede observar claramente que el mejor comportamiento, en cuanto a valores de precisión, lo presenta la curva perteneciente a la consulta expandida mediante la combinación de operadores *OR-AND* haciendo uso del PU, pues hasta un 17.8 % de documentos relevantes recuperados (Recall) presentan valores de precisión mucho más significativos puesto que son los primeros documentos recuperados, superando a los valores de precisión pertenecientes a las demás curvas de precisión. También se muestra que la precisión de la consulta de expansión con *OR-AND* sin hacer uso del PU está por debajo de la consulta sin expansión siendo esta última mucho más precisa.

También se observa que los valores más bajos de precisión pertenecen a las consultas de expansión con el operador *OR* haciendo uso del PU seguida de la consulta expandida con operadores *OR*, lo que indica que la precisión de los documentos recuperados hasta un 17.8 % de recall, es significativamente inferior a las demás consultas.

- **Identificador de Consulta:** 28

Consulta sin Expansión: palliation of cancer patients by using drugs, x-ray, physical phenomena or properties.

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
2	1	0,025641026	0,5
4	2	0,051282051	0,5
7	3	0,076923077	0,42857143
11	4	0,102564103	0,36363636

14	5	0,128205128	0,35714286
17	6	0,153846154	0,35294118
19	7	0,179487179	0,36842105

Tabla 34. Cálculo de Precisión y Recall Consulta 28 - Sin expansión

Precisión Promedio: 0,41010184

Consulta Expandida con operadores OR AND (sin PU): palliation OR cancer OR malignant neoplastic disease OR patients OR patient OR using OR exploitation AND drugs OR drug AND x OR ray OR beam of light OR phenomena OR phenomenon OR properties OR property OR ("Physical Phenomena or Properties" OR Physical) OR ("Physical Phenomena or Properties")) OR (X-Ray) OR Electromagnetic Radiation OR Ionizing Radiation OR Properties or Attributes.

Tipos de conceptos identificados:

CS → X-Ray
CC → Physical Phenomena or Properties CSI → Physical
CSE → Electromagnetic Radiation, Ionizing Radiation
T → palliation
(T → cancer, SIN → malignant neoplastic disease)
(T → patients, SIN → patient)
(T → using, SIN → exploitation)
(T → drugs, SIN → drug)
(T → ray, SIN → beam of light)
(T → phenomena, SIN → phenomenon)
(T → properties, SIN → property)

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
1	1	0,025641026	1
3	2	0,051282051	0,66666667
5	3	0,076923077	0,6
8	4	0,102564103	0,5
9	5	0,128205128	0,55555556
13	6	0,153846154	0,46153846
16	7	0,179487179	0,4375
20	8	0,205128205	0,4

Tabla 35. Cálculo de Precisión y Recall Consulta 28 - Operadores OR AND (sin PU)

Precisión Promedio: 0,577657585

Consulta Expandida con operadores OR AND (con PU): palliation OR cancer OR malignant neoplastic disease OR patients OR patient OR using OR exploitation AND drugs OR drug AND x OR ray OR beam of light OR phenomena OR phenomenon OR properties OR property OR ("Physical Phenomena or Properties" OR Physical) OR ("Physical Phenomena or Properties")) OR (X-Ray) OR Ionizing Radiation OR Properties or Attributes.

Tipos de conceptos identificados:

CS → X-Ray
CC → Physical Phenomena or Properties CSI → Physical
CSE → Electromagnetic Radiation, Ionizing Radiation
T → palliation
(T → cancer, SIN → malignant neoplastic disease)
(T → patients, SIN → patient)
(T → using, SIN → exploitation)
(T → drugs, SIN → drug)
(T → ray, SIN → beam of light)
(T → phenomena, SIN → phenomenon)
(T → properties, SIN → property)

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
1	1	0,025641026	1
2	2	0,051282051	1
3	3	0,076923077	1
5	4	0,102564103	0,8
7	5	0,128205128	0,71428571
9	6	0,153846154	0,66666667
12	7	0,179487179	0,58333333
15	8	0,205128205	0,53333333
17	9	0,230769231	0,52941176
19	10	0,256410256	0,52631579

Tabla 36. Cálculo de Precisión y Recall Consulta 28 - Operadores OR AND (con PU)

Precisión Promedio: 0,73533466

Consulta Expandida con operadores OR (sin PU): palliation OR cancer OR malignant neoplastic disease OR patients OR patient OR using OR exploitation OR drugs OR drug OR x OR ten OR ray OR beam of light OR phenomena OR phenomenon OR properties OR property OR ("Physical Phenomena or Properties" OR Physical) OR ("Physical Phenomena or Properties")) OR (X-Ray) OR Electromagnetic Radiation OR Ionizing Radiation OR Properties or Attributes.

Tipos de conceptos identificados:

CS → X-Ray
CC → Physical Phenomena or Properties CSI → Physical
CSE → Electromagnetic Radiation, Ionizing Radiation
T → palliation
(T → cancer, SIN → malignant neoplastic disease)
(T → patients, SIN → patient)
(T → using, SIN → exploitation)
(T → drugs, SIN → drug)
(T → ray, SIN → beam of light)
(T → phenomena, SIN → phenomenon)
(T → properties, SIN → property)

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
2	1	0,025641026	0,5
5	2	0,051282051	0,4
8	3	0,076923077	0,375
11	4	0,102564103	0,36363636
13	5	0,128205128	0,38461538
16	6	0,153846154	0,375
20	7	0,179487179	0,35

Tabla 37. Cálculo de Precisión y Recall Consulta 28 - Operador OR (sin PU)

Precisión Promedio: 0,392607393

Consulta Expandida con operadores OR (con PU): palliation OR cancer OR malignant neoplastic disease OR patients OR patient OR using OR exploitation OR drugs OR drug OR x OR ten OR ray OR beam of light OR phenomena OR phenomenon OR properties OR property OR ("Physical Phenomena or Properties" OR Physical) OR ("Physical Phenomena or Properties") OR (X-Ray) OR Ionizing Radiation OR Properties or Attributes.

Tipos de conceptos identificados:

CS → X-Ray
CC → Physical Phenomena or Properties CSI → Physical
CSE → Electromagnetic Radiation, Ionizing Radiation
T → palliation
(T → cancer, SIN → malignant neoplastic disease)
(T → patients, SIN → patient)
(T → using, SIN → exploitation)
(T → drugs, SIN → drug)
(T → ray, SIN → beam of light)
(T → phenomena, SIN → phenomenon)
(T → properties, SIN → property)

N° de documentos recuperados	N° Documentos Relevantes para el usuario	Recall	Precisión
1	1	0,025641026	1
5	2	0,051282051	0,4
8	3	0,076923077	0,375
11	4	0,102564103	0,36363636
14	5	0,128205128	0,35714286
16	6	0,153846154	0,375
20	7	0,179487179	0,35

Tabla 38. Cálculo de Precisión y Recall Consulta 28 - Operador OR (con PU)

Precisión Promedio: 0,460111317

RECALL	PRECISION				
	Expansión con operadores (OR AND) y PU	Expansión con operadores (OR AND)	Consulta sin Expansión	Expansión con operadores (OR) y PU	Expansión con operadores (OR)
0,02564103	1	1	0,5	1	0,5
0,05128205	1	0,666666667	0,5	0,4	0,4
0,07692308	1	0,6	0,428571429	0,375	0,375
0,1025641	0,8	0,5	0,363636364	0,363636364	0,363636364
0,12820513	0,714285714	0,555555556	0,357142857	0,357142857	0,384615385
0,15384615	0,666666667	0,461538462	0,352941176	0,375	0,375
0,17948718	0,583333333	0,4375	0,368421053	0,35	0,35
0,20512821	0,533333333	0,4			
0,23076923	0,529411765				

Tabla 39. Consulta 28 - Comparación curva Precisión-Recall de la consulta sin expansión y la expansión de consulta con operadores OR y operadores OR-AND evaluadas con y sin PU.

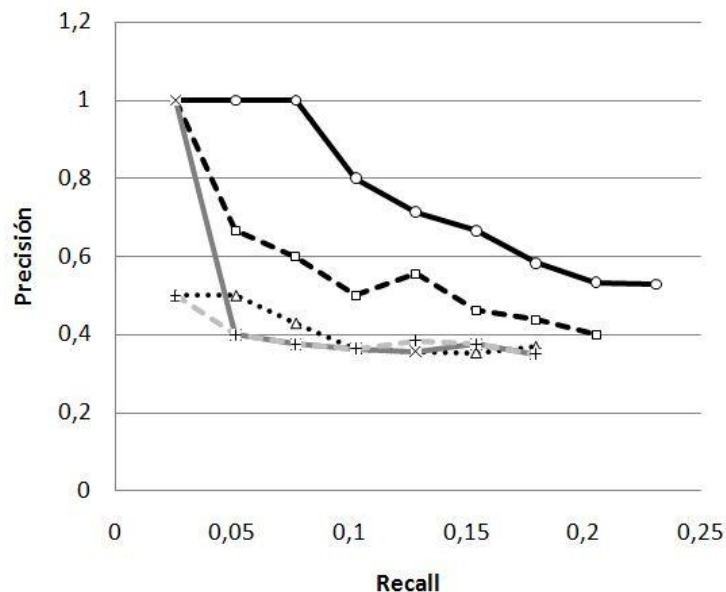


Figura 8. Curvas Precisión-Recuerdo para la tercera consulta.

En la gráfica anterior mostrada en la Figura 8, se puede observar claramente que el mejor comportamiento, en cuanto a valores de precisión, lo presenta la curva perteneciente a la consulta expandida mediante la combinación de operadores *OR-AND* haciendo uso del PU, la diferencia respecto a las demás curvas es muy significativa y además, esta curva de precisión-recall presenta valores de 1 en la precisión hasta alcanzar un 7.7% en el nivel de recall, lo que indica que se han obtenido varios documentos relevantes en las primeras posiciones del ranking que es donde usualmente se centra la atención del usuario [10]. A medida que se recuperan documentos, esta curva se mantiene por encima de las demás y sus valores de precisión se ubican en el rango de 0.8 – 0.52 alcanzando finalmente un nivel máximo de recall del 23.07%.

Vale la pena destacar que la segunda curva con mejor comportamiento es la que corresponde a la consulta expandida con los mismos operadores *OR-AND* pero sin

hacer uso del PU, esta curva supera también a la de la consulta sin expansión, que se ha establecido como la curva de referencia para realizar esta comparación. Lo anterior refleja el buen comportamiento en cuanto a recuperación de documentos relevantes al aplicar la expansión de consulta a través de la combinación de los operadores *OR-AND* además de la información que brinda el PU.

Respecto a las demás curvas, se puede concluir que su comportamiento no es el deseable comparado con el de las anteriores dos, si bien la curva correspondiente a la consulta expandida con el operador *OR* y el uso del PU alcanza el valor inicial 1 en la precisión, este decae muy rápidamente hasta alcanzar el valor 0.4, lo cual no es bueno teniendo en cuenta que los valores altos de precisión son obtenidos al localizar los documentos relevantes en las primeras posiciones del ranking, de este modo, los documentos relevantes se están situando en posiciones intermedias y de una forma alternada, es decir, separados por varios documentos no relevantes. Por último, la curva de la consulta expandida mediante el operador *OR* sin PU, posee un comportamiento similar al de la consulta sin expansión, tomando valores medios de precisión en los niveles iniciales de recall, por ende sólo se alcanza a recuperar una pequeña cantidad de documentos relevantes en las primeras posiciones del ranking.

En la se presentan los resultados de precisión promedio para cada una de las 15 consultas seleccionadas.

Id Consulta	Consulta	Formatos de Texto de Consulta	No. Documentos Relevantes Recuperados	Precisión Promedio
2	the relationship of blood and cerebrospinal fluid oxygen concentrations or partial pressures. a method of interest is polarography	Consulta sin Expansión	6	0,511
		Expansión con operadores (OR AND)	6	0,536
		Expansión con operadores (OR)	4	0,496
		Expansión con operadores (OR AND) y PU	7	0,598
		Expansión con operadores (OR) y PU	5	0,505
5	the crossing of fatty acids through the placental barrier. Normal fatty acid levels in placenta and fetus	Consulta sin Expansión	6	0,501
		Expansión con operadores (OR AND)	7	0,579
		Expansión con operadores (OR)	4	0,423
		Expansión con operadores (OR AND) y PU	9	0,611
		Expansión con operadores (OR) y PU	6	0,543
6	ventricular septal defect occurring in association with aortic regurgitation	Consulta sin Expansión	8	0,572
		Expansión con operadores (OR AND)	7	0,524
		Expansión con operadores (OR)	4	0,456
		Expansión con operadores (OR AND) y PU	7	0,524
		Expansión con operadores (OR) y PU	5	0,487
7	radioisotopes in heart scanning. mainly used in diagnosis of pericardial effusions. also used to study tumors, heart enlargement, aneurysms and pericardial thickening.	Consulta sin Expansión	9	0,578
		Expansión con operadores (OR AND)	8	0,556
		Expansión con operadores (OR)	7	0,528
		Expansión con operadores (OR AND) y PU	11	0,602

	technetium, rihsa, radioactive hippurate, cholegraffin are used	Expansión con operadores (OR) y PU	7	0,412
8	the effects of drugs on the bone marrow of man and animals, specifically the effect of pesticides. also, the significance of bone marrow changes	Consulta sin Expansión	6	0,566
		Expansión con operadores (OR AND)	6	0,562
		Expansión con operadores (OR)	5	0,432
		Expansión con operadores (OR AND) y PU	8	0,595
		Expansión con operadores (OR) y PU	5	0,432
9	the use of induced hypothermia in heart surgery, neurosurgery, head injuries and infectious diseases	Consulta sin Expansión	6	0,523
		Expansión con operadores (OR AND)	9	0,595
		Expansión con operadores (OR)	6	0,523
		Expansión con operadores (OR AND) y PU	7	0,542
		Expansión con operadores (OR) y PU	6	0,523
12	effect of azathioprine on systemic lupus erythematosus, particularly in regard to renal lesions	Consulta sin Expansión	5	0,436
		Expansión con operadores (OR AND)	9	0,586
		Expansión con operadores (OR)	7	0,512
		Expansión con operadores (OR AND) y PU	8	0,535
		Expansión con operadores (OR) y PU	5	0,417
14	renal amyloidosis as a complication of tuberculosis and the effects of steroids on this condition. only the terms kidney diseases and nephrotic syndrome were selected by the requester. prednisone and prednisolone are the only steroids of interest	Consulta sin Expansión	5	0,467
		Expansión con operadores (OR AND)	7	0,515
		Expansión con operadores (OR)	3	0,399
		Expansión con operadores (OR AND) y PU	8	0,546
		Expansión con operadores (OR) y PU	4	0,456
17	nickel in nutrition: requirements for methods for analysis; relation with enzyme systems; toxicity of, in humans and laboratory animals; deficiency signs and symptoms; level in various foodstuffs; level in blood and tissues	Consulta sin Expansión	7	0,496
		Expansión con operadores (OR AND)	6	0,450
		Expansión con operadores (OR)	3	0,321
		Expansión con operadores (OR AND) y PU	7	0,512
		Expansión con operadores (OR) y PU	5	0,357
19	excretion of phosphate or pyrophosphate in the urine or the effect of parathyroid hormone on kidney	Consulta sin Expansión	6	0,498
		Expansión con operadores (OR AND)	5	0,456
		Expansión con operadores (OR)	5	0,456
		Expansión con operadores (OR AND) y PU	5	0,456
		Expansión con operadores (OR) y PU	8	0,516
22	mycoplasma (infection or presence) in embryo, fetus, newborn infant or animal, or in pregnancy, gynecologic diseases, or as related to chromosomes or chromosome abnormaliti, or microanatomy	Consulta sin Expansión	8	0,512
		Expansión con operadores (OR AND)	8	0,455
		Expansión con operadores (OR)	5	0,342
		Expansión con operadores (OR AND) y PU	9	0,649
		Expansión con operadores (OR) y PU	6	0,376
24	compensatory renal hypertrophy stimulus resulting in mass increase (hypertrophy) and cell proliferation (hyperplasia) in the remaining kidney following unilateral nephrectomy in mammals	Consulta sin Expansión	5	0,471
		Expansión con operadores (OR AND)	6	0,511
		Expansión con operadores (OR)	3	0,389
		Expansión con operadores (OR AND) y PU	9	0,578
		Expansión con operadores (OR) y PU	3	0,389
	method for experimental	Consulta sin Expansión	10	0,562

26	production of and known cause of hydrocephalus in animal and human	Expansión con operadores (OR AND)	9	0.483
		Expansión con operadores (OR)	7	0.415
		Expansión con operadores (OR AND) y PU	12	0.579
		Expansión con operadores (OR) y PU	8	0.432
28	palliation of cancer patients by using drugs, x-ray, physical phenomena or properties	Consulta sin Expansión	7	0,410
		Expansión con operadores (OR AND)	8	0,577
		Expansión con operadores (OR)	7	0,392
		Expansión con operadores (OR AND) y PU	10	0,735
		Expansión con operadores (OR) y PU	7	0,460
30	hemophilia and christmas disease, especially in regard to the specific complication of pseudotumor formation (occurrence, pathogenesis, treatment, prognosis)	Consulta sin Expansión	7	0,512
		Expansión con operadores (OR AND)	7	0,530
		Expansión con operadores (OR)	7	0,535
		Expansión con operadores (OR AND) y PU	8	0,569
		Expansión con operadores (OR) y PU	7	0,500

Tabla 40. Precisión promedio de los cinco formatos de texto aplicados a 15 consultas de usuario de la colección MED.

En la se muestra el índice MAP calculado a partir de la precisión promedio presentada en la Tabla 40.

Id Consulta	Consulta sin Expansión	Expansión con operadores (OR AND)	Expansión con operadores (OR)	Expansión con operadores (OR AND) y PU	Expansión con operadores (OR) y PU
2	0,511	0,536	0,496	0,598	0,505
5	0,501	0,579	0,423	0,611	0,543
6	0,572	0,524	0,456	0,524	0,487
7	0,578	0,556	0,528	0,602	0,412
8	0,566	0,562	0,432	0,595	0,432
9	0,523	0,595	0,523	0,542	0,523
12	0,436	0,586	0,512	0,535	0,417
14	0,467	0,515	0,399	0,546	0,456
17	0,496	0,45	0,321	0,512	0,357
19	0,498	0,456	0,456	0,456	0,516
22	0,512	0,455	0,342	0,649	0,376
24	0,471	0,511	0,389	0,578	0,389
26	0,562	0,483	0,415	0,579	0,432
28	0,410	0,577	0,392	0,735	0,46
30	0,512	0,53	0,535	0,569	0,5
Índice MAP	0,503786	0,53085714	0,44314286	0,57307692	0,45138462

Tabla 41. Índice MAP de los cinco formatos de texto aplicados a 15 consultas de usuario de la colección MED.

6.1.2 Precisión en los K primeros resultados

Para calcular la precisión en los K primeros resultados de búsqueda se han tomado tres consultas que se muestran en la Tabla 42.

# Consulta	Texto de Consulta
1	Breast cancer treatment
2	Techniques of Cancer Control
3	Statistics of gastric Carcinoma of infant mortality in Africa

Tabla 42. Consultas utilizadas para calcular la medida Precision At K

6.1.2.1 Precision At k - MSEC Web Search

En la Tabla 43, Tabla 45 y Tabla 47 se muestran los juicios de relevancia de 15 observadores o jueces (R = Relevante, I = Irrelevante). Estos resultados son retornados por MSEC Web Search y se toman los 10 primeros resultados (k=10) que se muestran al usuario. En la Tabla 44, Tabla 46 y Tabla 48 se muestran los totales pertenecientes a cada categoría (relevante e irrelevante), además el porcentaje de exactitud (% Exa) y la Precisión en cada nivel de k (% Pre At-k).

DOCUMENTO	OBSERVADORES (USUARIOS)														
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
1	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
2	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
3	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
4	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
5	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
6	R	R	I	R	R	R	R	R	R	R	R	R	R	R	R
7	I	R	R	R	R	R	R	I	R	R	R	R	R	R	R
8	I	R	I	R	R	R	R	R	R	R	R	R	I	R	R
9	I	I	I	R	R	R	R	I	R	R	R	R	I	R	R
10	I	R	I	R	I	R	I	I	R	I	R	I	I	R	R

Tabla 43. Apreciación de los observadores en k=10 resultados (MSEC Web Search – Consulta 1)

Documento	Total: R	Total: I	% Exa	% Pre At-k
1	15	0	100	100
2	15	0	100	100
3	15	0	100	100
4	15	0	100	100
5	15	0	100	100
6	14	0	93,33	98,88
7	13	2	86,66	97,92
8	12	3	80	97,1
9	10	5	66,66	95,61
10	7	8	46,66	93,61
Precisión Promedio				0,98317196

Tabla 44. Exactitud y Precisión en cada nivel de k (MSEC Web Search – Consulta 1)

DOCUMENTO	OBSERVADORES (USUARIOS)														
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
1	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
2	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
3	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
4	R	I	R	R	R	R	I	R	R	R	R	R	R	R	R
5	R	R	R	I	R	R	I	R	R	R	R	R	R	R	R
6	R	R	I	I	R	R	R	I	R	R	R	R	R	R	I
7	R	I	R	I	R	R	I	I	R	R	R	R	I	R	R
8	R	R	R	I	R	R	I	R	R	R	R	I	R	R	I
9	R	I	R	I	R	R	I	R	R	I	R	I	R	I	I
10	R	I	R	I	R	R	I	R	I	R	I	I	R	R	R

Tabla 45. Apreciación de los observadores en k=10 resultados (MSEC Web Search – Consulta 2)

Documento	Total: R	Total: I	% Exa	% Pre At-k
1	15	0	1	1
2	15	0	1	1
3	15	0	1	1
4	13	2	0,86666667	0,96666667
5	13	2	0,86666667	0,96666667
6	11	4	0,73333333	0,94444444
7	10	5	0,66666667	0,93492063
8	11	4	0,73333333	0,94325397
9	8	7	0,53333333	0,92103175
10	9	6	0,6	0,92769841
Precisión Promedio				0,96046825

Tabla 46. Exactitud y Precisión en cada nivel de k (MSEC Web Search – Consulta 2)

DOCUMENTO	OBSERVADORES (USUARIOS)														
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
1	R	R	R	R	R	R	R	R	I	R	R	R	R	R	R
2	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
3	R	R	R	R	R	I	R	R	R	R	I	R	R	R	R
4	R	I	R	I	R	R	R	R	R	R	R	R	R	R	R
5	R	R	R	R	I	R	I	R	R	R	R	R	R	R	R
6	R	R	R	R	I	I	R	R	I	R	R	R	R	R	R
7	I	R	I	I	R	R	R	I	I	R	R	I	R	R	R
8	R	I	I	R	R	I	I	R	R	I	R	R	I	I	R
9	R	I	I	R	I	R	R	I	R	R	R	R	I	R	R
10	I	I	I	R	I	I	I	I	R	I	R	R	R	R	R

Tabla 47. Apreciación de los observadores en k=10 resultados (MSEC Web Search – Consulta 3)

Documento	Total: R	Total: I	% Exa	% Pre At-k
1	14	1	0,93333333	0,93333333
2	15	0	1	0,96666667
3	13	2	0,86666667	0,92222222
4	13	2	0,86666667	0,92222222
5	13	2	0,86666667	0,92222222
6	12	3	0,8	0,91111111
7	9	6	0,6	0,88253968
8	8	7	0,53333333	0,87420635
9	10	5	0,66666667	0,88902116
10	7	8	0,46666667	0,86902116
Precisión Promedio				0,90925661

Tabla 48. Exactitud y Precisión en cada nivel de k (MSEC Web Search – Consulta 3)

6.1.2.2 Precision At k - Google

En la Tabla 49, Tabla 51 y Tabla 53 se muestran los juicios de relevancia de 15 observadores o jueces (R = Relevante, I = Irrelevante). Estos resultados son retornados por Google y se toman los 10 primeros resultados (k=10) que se muestran al usuario. En la Tabla 50, Tabla 52 y Tabla 54 se muestran los totales pertenecientes a cada categoría (relevante e irrelevante), además el porcentaje de exactitud (% Exa) y la Precisión en cada nivel de k (% Pre At-k).

DOCUMENTO	OBSERVADORES (USUARIOS)														
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
1	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
2	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
3	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
4	R	R	R	I	R	R	R	R	R	R	R	R	R	R	R
5	R	R	R	I	R	R	R	R	R	R	R	R	R	R	R
6	R	R	R	I	R	I	I	R	R	R	R	R	R	R	R
7	R	I	R	I	I	R	I	I	R	I	R	R	R	R	R
8	I	R	I	I	I	R	I	I	I	I	R	R	R	R	R
9	I	I	I	R	I	R	I	I	I	I	R	I	I	I	R
10	I	I	I	I	I	I	I	I	R	I	R	I	I	I	I

Tabla 49. Apreciación de los observadores en k=10 resultados (Google – Consulta 1)

Documento	Total: R	Total: I	% Exa	% Pre At-k
1	15	0	100	100
2	15	0	100	100
3	15	0	100	100
4	14	1	93,33	98,33
5	14	1	93,33	98,33
6	12	3	80	96,11
7	9	6	60	93,25
8	7	8	46,66	91,58
9	4	11	26,66	89,36
10	2	13	13,33	88,02
Precisión Promedio				0,95501587

Tabla 50. Exactitud y Precisión en cada nivel de k (Google – Consulta 1)

DOCUMENTO	OBSERVADORES (USUARIOS)														
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
1	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
2	R	R	I	R	R	I	R	R	R	R	R	R	R	R	R
3	R	I	R	I	R	I	R	R	R	R	R	R	R	R	R
4	R	R	R	R	R	R	R	I	R	I	I	R	I	R	R
5	R	R	I	I	R	R	I	R	I	R	I	R	R	I	R
6	R	R	I	R	I	R	I	I	R	I	R	R	R	I	I
7	R	R	R	R	R	R	R	I	R	R	I	R	R	I	R
8	R	R	I	I	I	R	I	R	R	I	R	R	I	I	I
9	R	R	I	I	R	I	R	R	I	I	I	I	I	I	I
10	I	I	I	I	R	I	R	I	I	R	I	R	R	I	I

Tabla 51. Apreciación de los observadores en k=10 resultados (Google – Consulta 2)

Documento	Total: R	Total: I	% Exa	% Pre At-k
1	15	0	1	1
2	13	2	0,86666667	0,93333333
3	12	3	0,8	0,91111111
4	11	4	0,73333333	0,89444444
5	9	6	0,6	0,86777778
6	8	7	0,53333333	0,85666667

7	12	3	0,8	0,8947619
8	7	8	0,46666667	0,85309524
9	5	10	0,33333333	0,83828042
10	5	10	0,33333333	0,83828042
Precisión Promedio				0,88877513

Tabla 52. Exactitud y Precisión en cada nivel de k (Google – Consulta 2)

DOCUMENTO	OBSERVADORES (USUARIOS)														
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
1	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
2	R	R	R	I	I	R	R	I	R	R	R	R	R	R	R
3	R	R	R	I	R	R	R	R	R	R	R	R	R	R	R
4	R	R	I	R	R	I	R	R	I	R	R	I	R	R	R
5	R	R	R	I	R	R	I	R	I	R	R	R	R	R	R
6	R	I	R	I	R	I	I	I	R	I	R	R	R	R	R
7	R	I	R	R	I	R	R	R	I	R	R	R	R	I	I
8	R	R	I	I	I	R	R	I	R	R	R	I	R	I	R
9	R	I	I	I	R	R	I	I	I	R	I	R	R	R	R
10	R	R	R	I	R	I	R	R	I	R	R	R	R	R	R

Tabla 53. Apreciación de los observadores en k=10 resultados (Google – Consulta 3)

Documento	Total: R	Total: I	% Exa	% Pre At-k
1	15	0	1	1
2	12	3	0,8	0,9
3	14	1	0,93333333	0,94444444
4	11	4	0,73333333	0,89444444
5	12	3	0,8	0,90777778
6	9	6	0,6	0,87444444
7	10	5	0,66666667	0,88396825
8	9	6	0,6	0,87563492
9	8	7	0,53333333	0,86822751
10	12	3	0,8	0,89489418
Precisión Promedio				0,9043836

Tabla 54. Exactitud y Precisión en cada nivel de k (Google – Consulta 3)

6.1.2.3 Precision At k - Yahoo!

En la Tabla 55, Tabla 57 y Tabla 59 se muestran los juicios de relevancia de 15 observadores o jueces (R = Relevante, I = Irrelevante). Estos resultados son retornados por Yahoo! y se toman los 10 primeros resultados (k=10) que se muestran al usuario. En la Tabla 56, Tabla 58 y Tabla 60 se muestran los totales pertenecientes a cada categoría (relevante e irrelevante), además el porcentaje de exactitud (% Exa) y la Precisión en cada nivel de k (% Pre At-k).

DOCUMENTO	OBSERVADORES (USUARIOS)														
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
1	R	R	I	R	R	R	I	R	R	R	R	R	R	R	R
2	R	R	R	I	R	R	R	R	R	R	R	R	R	R	R
3	R	I	R	I	R	R	R	R	R	R	R	R	R	R	R
4	I	R	I	R	R	I	R	R	R	R	I	R	R	R	R

5	R	R	I	R	R	I	R	R	R	R	I	R	I	I	R
6	I	R	I	R	I	I	R	R	R	R	R	R	I	R	R
7	I	R	I	R	I	I	I	R	I	R	I	R	I	R	R
8	I	I	I	R	I	I	I	R	I	R	I	R	I	R	I
9	I	I	I	R	I	R	I	I	I	R	I	R	R	I	I
10	I	I	I	I	I	I	I	I	I	I	I	I	R	I	I

Tabla 55. Apreciación de los observadores en k=10 resultados (Yahoo! – Consulta 1)

Documento	Total: R	Total: I	% Exa	% Pre At-k
1	13	2	86,66	86,66
2	14	1	93,33	89,99
3	13	2	86,66	87,77
4	11	4	73,33	84,43
5	10	5	66,66	83,10
6	10	5	66,66	83,10
7	7	8	46,66	80,24
8	5	10	33,33	78,57
9	5	10	33,33	78,57
10	1	14	6,66	75,90
Precisión Promedio			0,82846032	

Tabla 56. Exactitud y Precisión en cada nivel de k (Yahoo! – Consulta 1)

DOCUMENTO	OBSERVADORES (USUARIOS)														
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
1	R	I	R	R	R	I	R	R	R	R	R	R	R	R	R
2	R	R	R	R	R	R	I	R	R	R	R	R	R	R	R
3	R	I	R	R	I	R	R	R	I	R	R	R	R	R	R
4	R	R	I	R	R	R	I	R	R	I	R	I	R	R	R
5	R	R	R	I	R	R	I	I	R	R	R	R	R	R	R
6	R	R	R	R	I	R	R	R	R	R	R	R	R	R	R
7	R	R	I	I	I	R	R	I	R	I	R	R	R	R	R
8	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
9	R	R	I	R	I	R	I	R	R	I	R	R	R	R	R
10	R	I	R	R	I	R	R	R	R	R	R	I	R	I	I

Tabla 57. Apreciación de los observadores en k=10 resultados (Yahoo! – Consulta 2)

Documento	Total: R	Total: I	% Exa	% Pre At-k
1	13	2	0,86666667	0,86666667
2	14	1	0,93333333	0,9
3	12	3	0,8	0,85555556
4	11	4	0,73333333	0,83888889
5	12	3	0,8	0,85222222
6	14	1	0,93333333	0,87444444
7	10	5	0,66666667	0,83634921
8	15	0	1	0,87801587
9	11	4	0,73333333	0,84838624
10	10	5	0,66666667	0,84171958
Precisión Promedio			0,85922487	

Tabla 58. Exactitud y Precisión en cada nivel de k (Yahoo! – Consulta 2)

DOCUMENTO	OBSERVADORES (USUARIOS)														
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
1	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
2	R	R	R	I	R	R	R	R	R	R	R	R	R	R	R
3	R	I	R	R	R	I	I	R	R	R	R	R	R	R	R
4	R	R	R	I	R	R	R	I	R	R	I	R	R	R	R
5	R	I	R	I	R	I	R	I	I	R	I	R	R	R	R
6	R	I	I	R	I	R	R	R	I	R	R	R	R	R	I
7	R	I	I	I	R	I	R	R	I	I	R	I	I	R	I
8	R	I	I	R	I	I	I	R	I	R	I	I	R	R	R
9	I	I	R	I	R	R	I	I	R	I	I	R	I	R	R
10	I	I	I	I	I	I	R	R	I	I	I	I	I	I	R

Tabla 59. Apreciación de los observadores en k=10 resultados (Yahoo! – Consulta 3)

Documento	Total: R	Total: I	% Exa	% Pre At-k
1	15	0	1	1
2	14	1	0,93333333	0,96666667
3	12	3	0,8	0,92222222
4	12	3	0,8	0,92222222
5	9	6	0,6	0,88222222
6	10	5	0,66666667	0,89333333
7	6	9	0,4	0,8552381
8	7	8	0,46666667	0,86357143
9	7	8	0,46666667	0,86357143
10	3	12	0,2	0,83690476
Precisión Promedio			0,90059524	

Tabla 60. Exactitud y Precisión en cada nivel de k (Yahoo! – Consulta 3)

6.1.3 Estadísticas Kappa

Se utilizó el *índice Kappa de Fleiss* [11, 12] que trabaja con cualquier número de observadores que proporcionan grados categóricos (en este caso R = Relevante, N = No relevante), a un número fijo de documentos (en este caso de k=1 hasta 10 documentos recuperados de la web). En la Tabla 62, Tabla 64 y Tabla 66, se muestra el juicio de los 15 observadores en los primeros 10 documentos recuperados para cinco consultas las cuales se presentan en la Tabla 61. En la Tabla 63, Tabla 65 y Tabla 67 se observan los valores de Kappa para los primeros 10 documentos recuperados.

# Consulta	Texto de Consulta
1	Organs affected by gastric carcinoma
2	Biological testing of DNA modification
3	Human cell division process

Tabla 61. Consultas utilizadas para calcular las estadísticas Kappa

DOCUMENTO	OBSERVADORES (USUARIOS)														
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
1	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
2	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
3	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R

4	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
5	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
6	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
7	I	R	R	I	R	R	I	I	I	R	R	I	I	I	R
8	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
9	I	I	I	R	I	I	R	R	I	I	I	I	I	I	I
10	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R

Tabla 62. Apreciación de los observadores en k=10 resultados para la consulta 1

Documento	Total: R	Total: I	Pi
1	15	0	1
2	15	0	1
3	15	0	1
4	15	0	1
5	15	0	1
6	15	0	1
7	7	8	0,46666667
8	15	0	1
9	3	12	0,65714286
10	15	0	1
Totales	130	20	9,12380952
pi	0,86666667	0,13333333	
pi²	0,75111111	0,01777778	
P̄	0,91238095		
P̄e	0,76888889		
Kappa	0,62087912		

Tabla 63. Resultado del índice kappa en k=10 para la consulta 1

DOCUMENTO	OBSERVADORES (USUARIOS)														
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
1	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
2	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
3	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
4	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
5	I	I	I	R	I	I	I	R	I	R	I	R	I	I	I
6	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
7	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
8	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
9	R	I	I	I	R	I	R	I	I	R	I	I	I	I	I
10	I	R	I	I	I	I	I	I	I	I	I	R	I	I	I

Tabla 64. Apreciación de los observadores en k=10 resultados para la consulta 2

Documento	Total: R	Total: I	Pi
1	15	0	1
2	15	0	1
3	15	0	1
4	15	0	1
5	4	11	0,58095238
6	15	0	1
7	15	0	1
8	15	0	1

9	4	11	0,58095238
10	2	13	0,75238095
Totales	115	35	8,91428571
π	0,76666667	0,23333333	
π^2	0,58777778	0,05444444	
\bar{P}	0,89142857		
$\bar{P}e$	0,64222222		
Kappa	0,69653949		

Tabla 65. Resultado del índice kappa en k=10 para la consulta 2

DOCUMENTO	OBSERVADORES (USUARIOS)														
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15
1	R	R	R	I	R	R	R	R	R	R	I	R	R	R	I
2	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
3	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
4	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
5	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
6	I	R	I	I	I	R	R	I	R	R	I	I	I	I	R
7	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
8	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
9	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
10	I	R	I	I	I	R	I	I	R	I	I	I	I	I	R

Tabla 66. Apreciación de los observadores en k=10 resultados para la consulta 3

Documento	Total: R	Total: I	P_i
1	12	3	0,65714286
2	15	0	1
3	15	0	1
4	15	0	1
5	15	0	1
6	6	9	0,48571429
7	15	0	1
8	15	0	1
9	15	0	1
10	4	11	0,58095238
Totales	127	23	8,72380952
π	0,84666667	0,15333333	
π^2	0,71684444	0,02351111	
\bar{P}	0,87238095		
$\bar{P}e$	0,74035556		
Kappa	0,50848535		

Tabla 67. Resultado del índice kappa en k=10 para la consulta 3

Teniendo en cuenta los valores de las estadísticas Kappa Fleiss para los resultados de MSEC Web Search, se puede observar que la concordancia entre los jueces es buena ubicándose entre una moderada concordancia (valores entre 0,41 y 0,60) y una concordancia sustancial (valores entre 0,61 y 0,80) de acuerdo al estudio de Landis y Koch [13].

6.1.4 Índice MAP de GoPubMed y MSEC Web Search

En la se presenta la precisión promedio para cada una de las 15 consultas realizadas sobre el repositorio de documentos MEDLINE y en la se muestra el índice MAP calculado a partir de estas precisiones promedio.

Id Consulta	Consulta	SRI	Precisión Promedio
2	the relationship of blood and cerebrospinal fluid oxygen concentrations or partial pressures. a method of interest is polarography	GoPubMed	0,783
		MSEC Web Search	0,643
5	the crossing of fatty acids through the placental barrier. Normal fatty acid levels in placenta and fetus	GoPubMed	0,722
		MSEC Web Search	0,698
6	ventricular septal defect occurring in association with aortic regurgitation	GoPubMed	0,719
		MSEC Web Search	0,589
7	radioisotopes in heart scanning. mainly used in diagnosis of pericardial effusions. also used to study tumors, heart enlargement, aneurysms and pericardial thickening. technetium, rihsa, radioactive hippurate, cholegraffin are used	GoPubMed	0,623
		MSEC Web Search	0,623
8	the effects of drugs on the bone marrow of man and animals, specifically the effect of pesticides. also, the significance of bone marrow changes	GoPubMed	0,542
		MSEC Web Search	0,428
9	the use of induced hypothermia in heart surgery, neurosurgery, head injuries and infectious diseases	GoPubMed	0,756
		MSEC Web Search	0,611
12	effect of azathioprine on systemic lupus erythematosus, particularly in regard to renal lesions	GoPubMed	0,89
		MSEC Web Search	0,745
14	renal amyloidosis as a complication of tuberculosis and the effects of steroids on this condition. only the terms kidney diseases and nephrotic syndrome were selected by the requester. prednisone and prednisolone are the only steroids of interest	GoPubMed	0,751
		MSEC Web Search	0,703
17	nickel in nutrition: requirements for methods for analysis; relation with enzyme systems; toxicity of, in humans and laboratory animals; deficiency signs and symptoms; level in various foodstuffs; level in blood and tissues	GoPubMed	0,779
		MSEC Web Search	0,715
19	excretion of phosphate or pyrophosphate in the urine or the effect of parathyroid hormone on kidney	GoPubMed	0,658
		MSEC Web Search	0,634
22	mycoplasma (infection or presence) in embryo, fetus, newborn infant or animal, or in pregnancy, gynecologic diseases, or as related to chromosomes or chromosome abnormaliti, or microanatomy	GoPubMed	0,681
		MSEC Web Search	0,552
24	compensatory renal hypertrophy stimulus resulting in mass increase (hypertrophy) and cell proliferation (hyperplasia) in the remaining kidney following unilateral nephrectomy in mammals	GoPubMed	0,597
		MSEC Web Search	0,499
26	method for experimental production of and known cause of hydrocephalus in animal and human	GoPubMed	0,745
		MSEC Web Search	0,678
28	palliation of cancer patients by using drugs, x-ray, physical phenomena or properties	GoPubMed	0,688
		MSEC Web Search	0,581
30	hemophilia and christmas disease, especially in regard to the specific complication of pseudotumor formation (occurrence, pathogenesis, treatment, prognosis)	GoPubMed	0,735
		MSEC Web Search	0,68

Tabla 68. Precisión promedio de los resultados devueltos por GoPubMed y MSEC Web Search a través de 15 consultas sobre MEDLINE.

Id Consulta	GoPubMed	MSEC Web Search
2	0,783	0,643
5	0,722	0,698
6	0,719	0,589
7	0,623	0,623
8	0,542	0,428
9	0,756	0,611
12	0,89	0,745
14	0,751	0,703
17	0,779	0,715
19	0,658	0,634
22	0,681	0,552
24	0,597	0,499
26	0,745	0,678
28	0,688	0,581
30	0,735	0,68
Índice MAP	0,71126667	0,62526667

Tabla 69. Precisión promedio de los resultados devueltos por GoPubMed y MSEC Web Search a través de 15 consultas sobre MEDLINE.

7 ANEXO G

7.1 Manual de Usuario

Este manual de usuario ilustra la forma de utilizar la aplicación “MSEC Web Search” y comprende una especificación del prototipo orientada hacia las búsquedas en la Web, específicamente del dominio de la Oncología. A continuación se muestran las funcionalidades del prototipo “MSEC Web Search”.

7.1.1 Ingresar al sistema

El usuario ingresa al sistema mediante su Login y su Password, para ello se debe digitar la siguiente dirección en el explorador Web como se muestra en la Figura 9:

<http://prometeo.unicauca.edu.co/msec/ContentPages/Usuarios/InterfazBuscador.aspx>

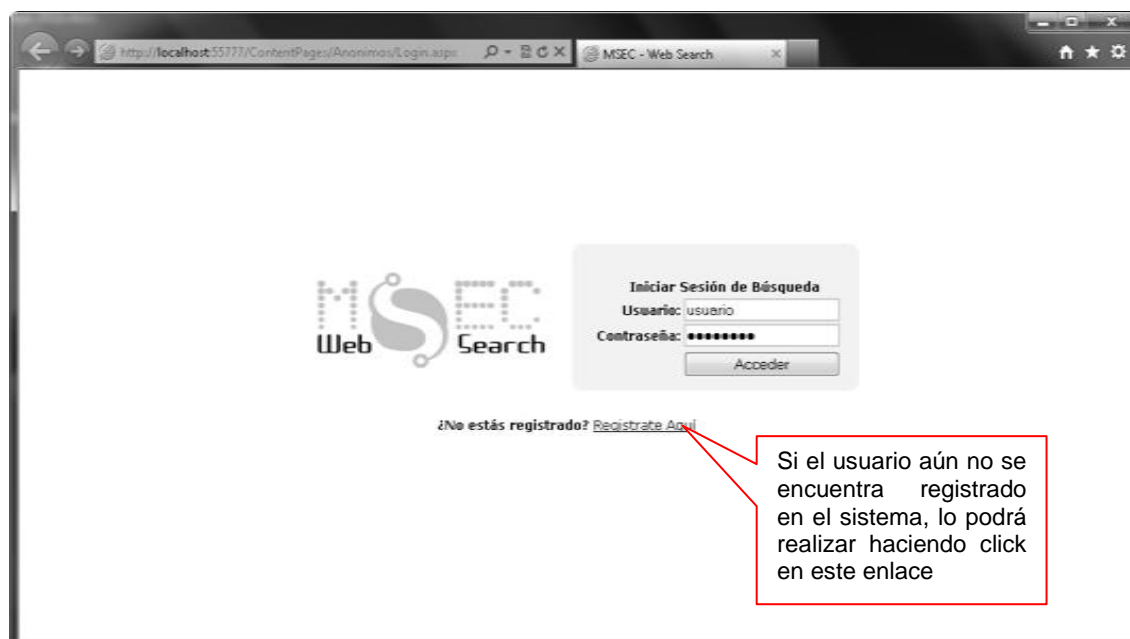


Figura 9. Interfaz de inicio de sesión.

7.1.2 Registrar Usuario

Con el objetivo de crear un Perfil de Usuario para realizar búsquedas Web, se deben registrar un Login y un Password tal como se muestra en la Figura 10.

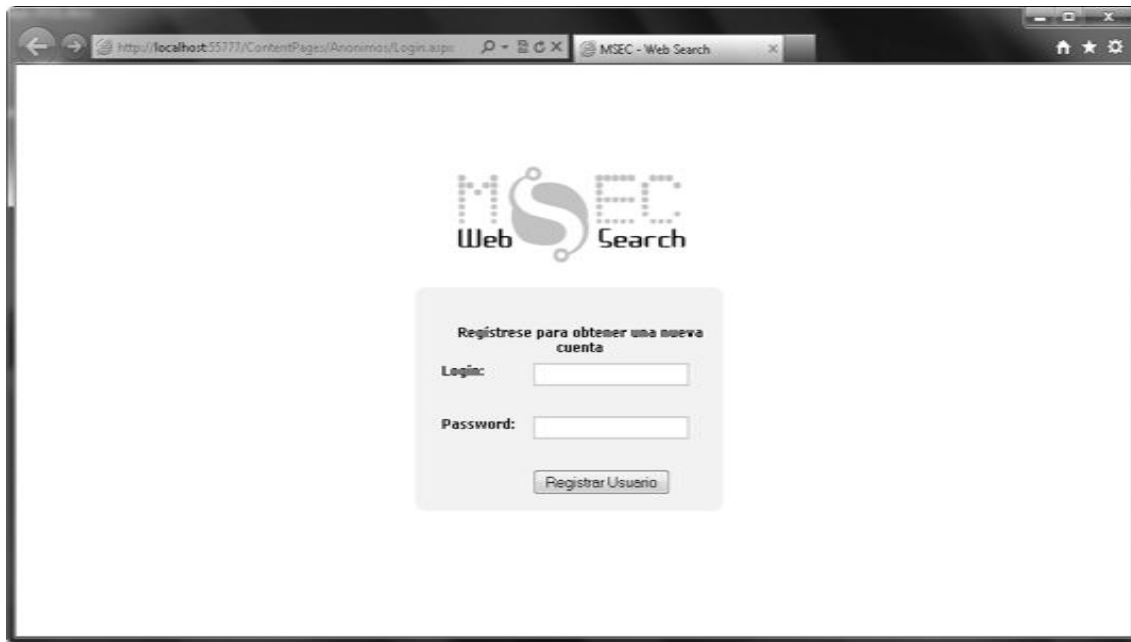


Figura 10. Interfaz Registrar Usuario

7.1.3 Búsqueda Web

Una vez que el usuario ha ingresado al sistema, puede realizar búsquedas en el dominio de la Oncología por medio de la interfaz de búsqueda mostrada en la Figura 11.

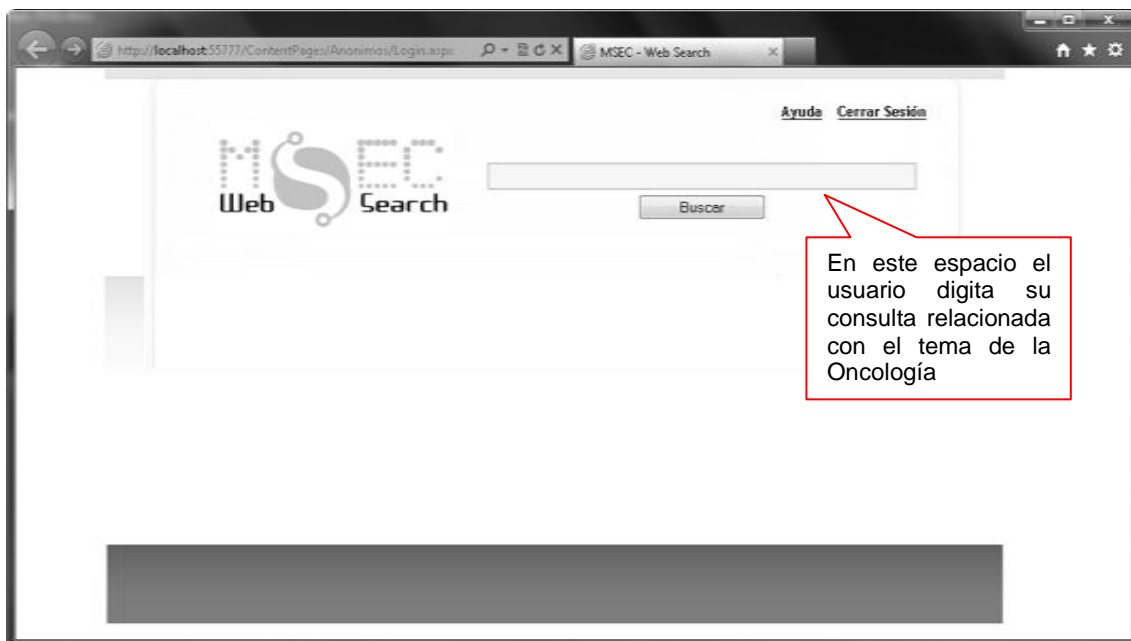


Figura 11. Interfaz de Búsqueda.

Cuando el usuario digita su consulta y presiona el botón “Buscar” se muestran los resultados de búsqueda como se puede ver en la Figura 12.

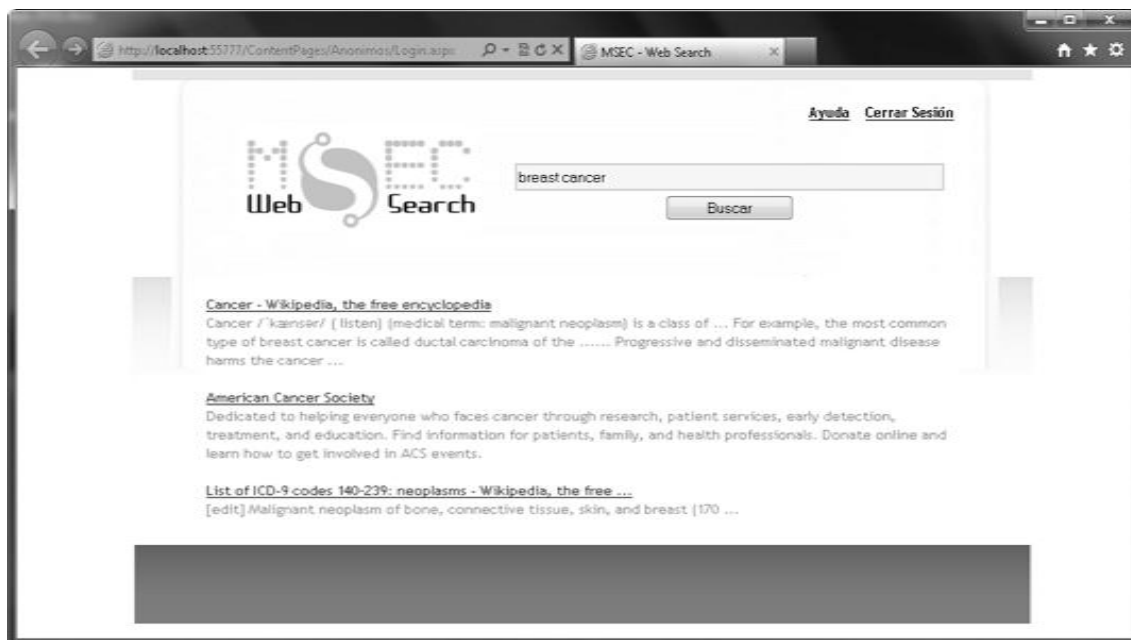


Figura 12. Resultados de Búsqueda

7.1.4 Obtener Ayuda

Para obtener ayuda sobre la aplicación y autores de la misma, puede ingresar al botón “Ayuda” en la parte superior derecha de la página como se muestra en la Figura 13.

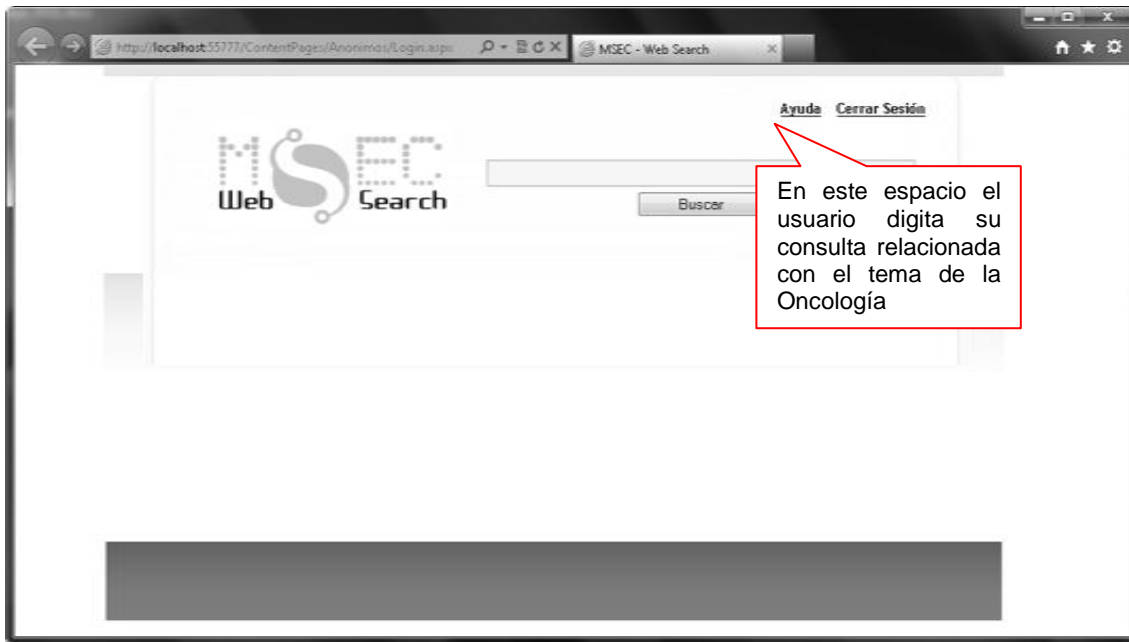


Figura 13. Ayuda de la Aplicación.

7.1.5 Cerrar Sesión

Para terminar la sesión de búsqueda y salir del sistema, puede dar click al botón “Cerrar Sesión” en la parte superior derecha de la página como se muestra en la Figura 14.

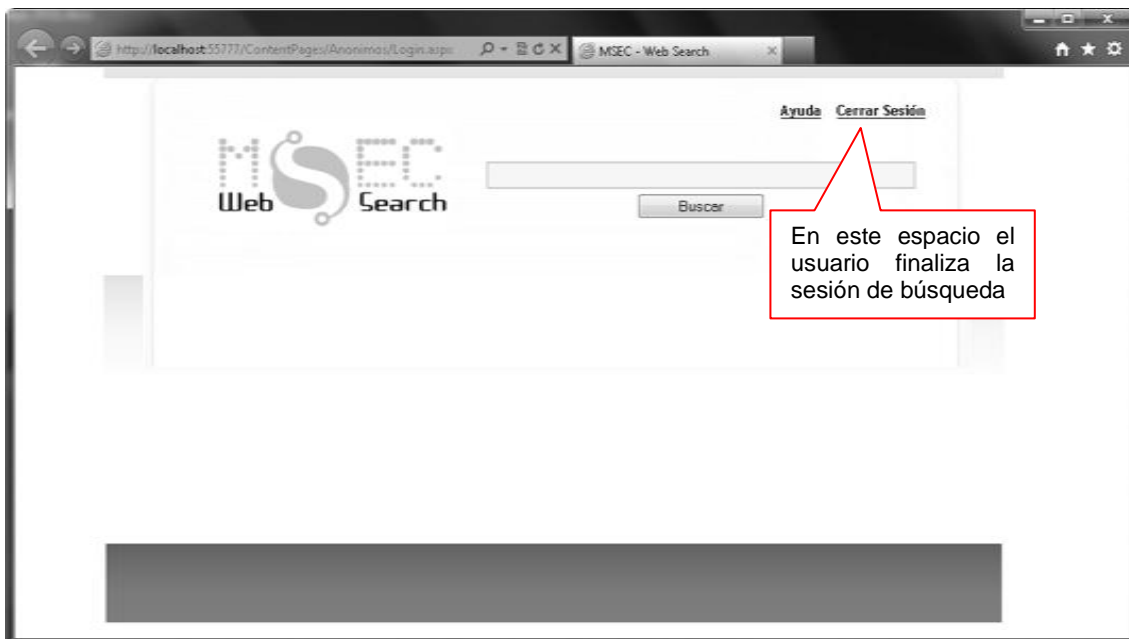


Figura 14. Cerrar Sesión.

8 REFERENCIAS

- [1] M. Grüninger and M. Uschold, "Ontologies and semantic integration," *Government report on the state of the art and future predictions for agent technology*, 2002.
- [2] S. Bechhofer, F. V. Harmelen, and J. Hendler, "OWL Web ontology language " *Recomendación W3C*, 2004.
- [3] Wikipedia, "Oncología", [en línea] <http://es.wikipedia.org/wiki/Oncolog%C3%ADa>. [Consulta: 07 de Mayo de 2011].
- [4] NCI, "National Cancer Institute. U.S. National Institutes of Health", [en línea] <http://www.cancer.gov>. [Consulta: 26 de Abril de 2011].
- [5] Manchester_University, "The University Of Manchester", [en línea] www.manchester.ac.uk/. [Consulta: 24 de Mayo de 2011].
- [6] OpenGALEN, "GRAIL Tutorial", [en línea] <http://www.opengalen.org/tutorials/grail/tutorial1.html>. [Consulta: 24 de Mayo de 2011].
- [7] P. Kruchten, "Planos Arquitectónicos: El Modelo de "4+1" Vistas de la Arquitectura del Software," *IEEE Software*, vol. 12, 1995.
- [8] MSDN_Microsoft, "¿Qué es un Patrón de Diseño?", [en línea] <http://msdn.microsoft.com/es-es/library/bb972240.aspx>. [Consulta: 13 de Mayo de 2011].
- [9] MEDLARS, "MEDical Literature Analysis and Retrieval System", [en línea] <http://www.uninet.edu/do/MEDLARS.html>. [Consulta: 26 de Abril de 2011].
- [10] N. Lahkar and S. K. Deka, "Impact of Query Operators on Web Search Engine Results : An Evaluative Study," in *2nd Convention PLANNER - 2004* Manipur Uni., Imphal, 2004.
- [11] F. Martínez, Propuesta y desarrollo de un modelo para la evaluación de la recuperación de información en Internet, in *Información y Documentación*. 2002, Universidad de Murcia: Murcia, España. p. 283.
- [12] F. Cacheda, V. Formoso, and V. Carneiro, Performance Analysis of Distributed Web Information Retrieval Systems. *Latin America Transactions, IEEE (Revista IEEE America Latina)*, 2007. 5(6): p. 479-485.
- [13] J. Landis and G. Koch, "The measurement of observer agreement for categorical data" in *Biometrics*. Vol. 33, 1977, pp. 159–174.