MIS0008
2015
Ej. 1

1323846

NAVEGADOR ONTOLÓGICO MATEMÁTICO

NOMAT

ERNESTO VARELA ARREGOCÉS

UNIVERSIDAD SIMÓN BOLIVAR

MAESTRÍA EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

BARRANQUILLA, 2015

UNIVERSIDAD SIMÓN BOLIVAR
BIBLIOTECA
INSTITUTO DE POSTGRADO

NAVEGADOR ONTOLÓGICO MATEMÁTICO

NOMAT

ERNESTO VARELA ARREGOCÉS

Trabajo presentado para optar el título de Magister en Ingeniería de Sistemas y computación

Asesores:

Phd. CARLOS FAJARDO TORO

Phd. PAOLA SÁNCHEZ SÁNCHEZ

UNIVERSIDAD SIMÓN BOLIVAR

MAESTRÍA EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

BARRANQUILLA, 2015

UNIVERSIDAD SIMON BOLIVAR
BIBLIOTECA
INSTITUTO DE POSTGRADO

DEDICATORIA

Este trabajo se lo dedico a Dios, por estar siempre presente en mi vida, a mi familia: a mi Madre quien me dio la vida y siempre me apoyó durante el proceso de mi formación académica hasta convertirme en profesional, a mi Padre por enseñarme el valor del trabajo, a mi hermana Rocío quien con su prudencia y sabiduría me apoyó y revisó este trabajo hasta sacar un escrito coherente, a mi hermana Gloria quien con su pregunta frecuente que casi llegó a enloquecerme: "¿Cómo va la tesis?", ¡finalmente logré culminarla Gloria !, a mi hermana Silvia y mis bellas sobrinas Laura y Andrea, gracias por sacarme de la rutina, a mis amigos: Mónica, Raúl, Armando, José Fernando, por siempre sacarme una sonrisa en medio de las dificultades, a mis profesores, asesores y a la Universidad, por educarme y brindarme sus conocimientos, para todos ellos…

$$r = a(1 + \cos\theta) \ \ldots$$

¡GRACIAS TOTALES!

TABLA DE CONTENIDO

	Página
RESUMEN	6
TERMINOLOGÍA BÁSICA RELACIONADA	8
LISTA DE TABLAS	12
LISTA DE GRÁFICOS	13
INTRODUCCIÓN	14
CAPÍTULO 1	19
1.1. Estado del Arte	19
1.2. Marco Teórico Conceptual	21
1.2.1. Web Semántica	21
1.2.2. Conceptos Relacionados con Web Semántica	23
1.2.3. Elementos de la Web Semántica	26
1.2.4. Servicios de la Web Semántica	28
1.2.5. Aplicaciones Técnicas de la Web Semántica	29
1.2.6. Ventajas y Desventajas de la Web Semántica	31
1.2.7. Ontologías	32
1.2.8. Contexto Definitorio (CD)	40
CAPÍTULO 2	51
2.1. Metodología	51
2.1.1. Tipo de Investigación	51
2.1.2. Diseño de Investigación	52
2.1.3. Técnicas de Procesamiento de Datos	53
CAPÍTULO 3	55
3.1. Descripción Técnica de NOMAT	55
3.2. Diagrama de Clases	55
3.3. Diccionario de Clases	57
3.4. Diseño del Sistema	63
3.4.1. Diseño de la Ontología	64
3.4.2. Diseño de la Base de Datos	65
3.4.3. Diseño de la Interfaz Gráfica	67
3.5. Implementación y Desarrollo	69
3.5.1. Interfaz de Usuario	69
3.5.2. Controles	73
3.5.3. Modelo	77
3.5.4. Librerías	79
3.6. Software de Apoyo	85
CAPÍTULO 4	87
4.1. Experimentación	87
4.2. Modelo de Búsqueda de NOMAT	92
4.2.1. Componente de consulta	93

INSTITUTO DE POSTGRADO

4.2.2. Componente de Captura de Contenidos 95
4.2.3. Componente de Filtrado 95
4.2.4. Componente de Recuperación 97
4.2.5. Componente de Visualización de la Información 97
4.2.6. Componente de Almacenamiento 98
4.3. Experimentando con NOMAT 100

CONCLUSIONES 103

RECOMENDACIONES 105

REFERENCIAS BIBLIOGRÁFICAS 107

RESUMEN

La consulta de información en internet es un aspecto que se ha ido desarrollando desde sus inicios y se han logrados muchos avances en este sentido, hay algoritmos de consulta de muchos tipos, que emplean indexación, análisis contextual, ontologías etc. Habitualmente las consultas que se hacen solo trabajan el texto, pero no se emplean ecuaciones por la complejidad que estas generan al tratar de escribirlas, basta con hacer una prueba sencilla en alguno de los motores de búsqueda comerciales conocidos para darse cuenta de ello. NOMAT es un primer prototipo de motor de Búsqueda de Expresiones matemáticas que usa una herramienta Ontológica que realiza este tipo de consulta, utiliza texto complementario y Látex como editor de la simbología empleada en este tipo de expresiones. Esta herramienta, puede buscar información en un Tesaurus y en internet directamente. Fue creada como herramienta de apoyo a investigaciones de tipo matemático.

Comparada con otros buscadores de internet y aunque estos pueden hacer la consulta directamente en Látex, NOMAT es el buscador que no exige el conocimiento de dicho lenguaje, sino que cuenta con la herramienta de edición y simplemente se puede escribir directamente la expresión y consultarla. Los

resultados mostrados fueron bastante precisos y brindan la base para una investigación futura en torno a emplear agentes inteligentes y a plantear la exigencia de que toda publicación de tipo matemático sea estandarizada en Látex.

TERMINOLOGÍA BÁSICA RELACIONADA

Látex: LaTeX es un sistema editor de documentos y lenguaje de marcas para el programa TeX. El término LaTeX hace referencia solo al lenguaje en que estos documentos son escritos, pero no al editor utilizado para escribirlos. Para crear un documento en LaTeX, un archivo .tex debe ser creado empleando un editor de textos. Cualquier editor puede ser utilizado para crear un documento LaTeX, pero existen también algunos editores específicos para trabajar con LaTeX. Es ampliamente usado por matemáticos, científicos, ingenieros, economistas, investigadores y estudiantes en general.

Tesaurus: El término tesauro tiene su origen etimológico en la palabra latina thesaurus, que, a su vez, tiene su origen en el vocablo griego thesaurós. En ambos casos, el significado era tesoro o repositorio de palabras.

Un tesauro es "un tipo de lenguaje documental que representa la estructuración conceptual de un determinado campo del conocimiento. El tesauro proporciona una organización semántica principalmente a través de la explicitación de las relaciones establecidas entre dichos conceptos y, eventualmente, a través de un significado restringido de los

términos que los representan. Los tesauros, para la Documentación y la Biblioteconomía, son por lo tanto una herramienta de control terminológico, puesto que la estructuración del tesauro es en base a conceptos. Los conceptos, son representados por términos seleccionados que evidencian ese control terminológico. Con dicho control se busca neutralizar la sinonimia y la polisemia, ambas características naturales de la lengua, que dificultan la precisión de la indización y recuperación de la información, ambas funciones básicas de los tesauros. Los tesauros son herramientas construidas para el auxilio tanto de los profesionales de la información como de los usuarios finales". (Arano, 2005)

Ontología: Se considera como una representación explícita y formal de una conceptualización compartida, que implica una perspectiva sobre cierta realidad y que se constituye en la estructura conceptual de una base de conocimiento. Su objetivo final, como se rescata de las definiciones estudiadas, es el compartir el conocimiento que representa. (Arano, 2005)

Navegador Web(Browser):Los navegadores web permiten navegar por Internet. Hay una variedad de opciones disponibles, por lo tanto puede elegir la que mejor satisfaga sus necesidades.

Un navegador web es una aplicación que encuentra y muestra páginas web. Coordina la comunicación entre la computadora y el servidor web donde "reside" un sitio web en particular.

Cuando se abre el navegador y se escribe una dirección web (URL) para un sitio web, este contacta ese servidor, solicita la página web que la persona pidió, y muestra la página en su computadora. El navegador traduce el código (escrito en un idioma tal como HTML o XML) para los diferentes elementos de la página (texto, imágenes, sonidos) a un formato apropiado y muestra la página resultante.

Contexto de una expresión matemática: Interpretar contextualmente una expresión de tipo matemático es un verdadero reto para los seres humanos, por ende para una máquina es un proceso mucho más complejo. Por ello se deben tener en cuenta signos, que indican operaciones o acciones a realizar o equivalencias; las letras o variables pueden representar valores. Aparte de lo anterior, están también los respectivos operadores específicos, tales como las derivadas, derivadas parciales o las integrales, entre otros; y todo esto conlleva a la comprensión de un lenguaje que puede variar su significado dependiendo del contexto. Una expresión matemática

UNIVERSIDAD SIMON BOLIVAR
BIBLIOTECA
INSTITUTO DE POSTGRADO

sola no tiene ningún sentido, si no conlleva un texto complementario que la contextualice.

LISTA DE TABLAS

Página

Tabla 1	Investigaciones Relacionadas con la búsqueda Contextual en Matemáticas Discriminada Cronológicamente.	20
Tabla 2	Los 7 Niveles de la Web Semántica	25
Tabla 3	Elementos de la Web Semántica	26
Tabla 4	Ventajas y Desventajas de la Web Semántica	31
Tabla 5	Algunos Conceptos de Ontología en Informática	34
Tabla 6	Terminología Relacionada con las Ontologías	36
Tabla 7	Propuesta de Contrastación del comportamiento de los Buscadores comerciales más Importantes en Internet a partir de Expresiones matemáticas.	54
Tabla 8	Clase Gestionontologia	58
Tabla 9	Método # 1 Gestionontologia	58
Tabla 10	Método # 2 Gestionontologia	59
Tabla 11	Clase GestionContenidos	59
Tabla 12	Método # 1 GestionContenidos	60
Tabla 13	Método # 2 GestionContenidos	60
Tabla 14	Método # 3 GestionContenidos	61
Tabla 15	Método # 4 GestionContenidos	61
Tabla 16	Clase Consultas	62
Tabla 17	Método # 1 Consultas	62
Tabla 18	Método # 2 Consultas	63
Tabla 19	Tipos de conversión entre Java y Java Script.	84
Tabla 20	Contrastación de Comportamientos de Consultas de Expresiones matemáticas en los buscadores de Internet	91
Tabla 21	Comportamiento de las Consultas editadas en LATEX hechas en NOMAT	100

LISTA DE GRÁFICOS

Página

Gráfico 1	Modelo de Arquitectura de la Web Semántica según World Wide Web Consortium	24
Gráfico 2	Icono de las páginas RSS	26
Grafico 3	Servicios de la Web Semántica	28
Gráfico 4	Ejemplo de Un Contexto Definitorio	44
Gráfico 5	Diagrama de Clases de NOMAT	56
Gráfico 6	Árbol Jerárquico de la Ontología Empleada en NOMAT	64
Gráfico 7	Diagrama Ontológico de NOMAT	65
Gráfico 8	Diagrama de la Base de Datos de NOMAT	66
Gráfico 9	Interfaz gráfica del Navegador NOMAT	67
Gráfico 10	Panel de Resultados de Consulta	68
Gráfico 11	Consulta en Google de la expresión 2x+3y=1	88
Grafico 12	Consulta en Yahoo de la expresión 2x+3y=1	88
Gráfico 13	Consulta en AOL de la expresión 2x+3y=1	89
Gráfico 14	Consulta en Excite de la expresión 2x+3y=1	90
Gráfico 15	Consulta en Bing de la expresión 2x+3y=1	93
Gráfico 16	Arquitectura de NOMAT	94
Gráfico 17	Consulta en NOMAT	98
Gráfico 18	Visualización de la Información en NOMAT	99
Gráfico 19	Visualización del Criterio de Búsqueda en NOMAT	99
Gráfico 20	Visualización del URL almacenada en NOMAT	101
Gráfico 21	Resultados de Consulta de la Ecuación Diferencial buscada	102
Gráfico 22	Resultados de la Consulta de la integral buscada	99

INTRODUCCIÓN

Desde los años 90 cuando se presenta oficialmente la World Wide Web ante el mundo, se evidenció la ausencia de los contextos en las búsquedas para que éstas fueran mucho más precisas a la hora de utilizar un navegador (comerciales y no comerciales), y aunque actualmente existen algunos avances en este sentido, aun se pierde tiempo valioso en el proceso, ya que la web fue basada en documentos para ser leída por humanos y no comprendida por las computadoras (Berners_lee Tim, 2002).

La necesidad de extraer información específica a partir de textos, se está convirtiendo en una acción frecuente para actividades relacionadas con el procesamiento del lenguaje natural, tales como: lexicografía computacional, semántica computacional, minería de textos, web semánticas y aprendizaje automático a partir de patrones lingüísticos, etc., de hecho, la inteligencia artificial ha sido el área que más ha trabajado en este sentido, porque ha tomado como núcleo principal la extracción terminológica y conceptual como base para el desarrollo de las tecnologías del lenguaje, de ahí que sea necesario contar con un adecuado sistema de cómputo para efectuar las consultas en forma automatizada.

En este mismo sentido, en la web se encuentran investigaciones de tipo matemático publicadas en bases de datos especializadas de consulta y en internet en general, usando herramientas de edición que son de uso exclusivo de expertos tales como, el MATHML® y el LATEX®, entre otras, que facilitan este proceso, debido a que la transcripción de muchas expresiones resulta complicada y el teclado habitual de las computadoras no cuenta con la simbología adecuada, o en su defecto las ecuaciones son tratadas como imágenes. Estos aspectos permiten que algunos navegadores de internet como Microsoft Explorer®, Google Chrome®, Safari® y Mozilla® entre otros navegadores comerciales, reconozcan dichas publicaciones, pero estos tampoco identifican el contexto semántico de lo publicado en la página, basta con tratar de introducir una expresión compleja para comprobarlo.

Siguiendo esta misma línea, existe herramienta de búsqueda de tipo científico y más orientado hacia las investigaciones en el campo de las matemáticas y la física llamado "MathSciNet" de la Sociedad Americana de Matemáticas que en efecto busca material científico con bastante precisión, pues cuenta con herramientas que ayudan a realizar la consulta de manera específica, sin embargo tampoco hace una búsqueda a partir de ecuaciones y el texto complementario.

Otra herramienta que hace lo mismo es “Wolfram Alpha”, que aunque da la sensación de resolver problemas, lo que en verdad hace es una consulta en su propio repositorio de información que es alimentado por un gran equipo de matemáticos que solucionan un sin número de ejercicios; y aunque en este caso se permite escribir las ecuaciones (no de forma precisa sino con comandos propios), esta herramienta no hace consulta en la web directamente.

Basado en la revisión bibliográfica y en la revisión de algunos de los navegadores mencionados se puede afirmar que no se ha desarrollado una técnica de consulta contextual de expresiones matemáticas en los navegadores de internet que precise los resultados de búsquedas especializadas.

Las situaciones antes descritas permitieron formular la siguiente pregunta de investigación: ¿cómo desarrollar un prototipo de navegador web de tipo matemático que facilite la consulta contextual de expresiones matemáticas?

Como respuesta a esta problemática surge el primer prototipo llamado: **NOMAT Navegador Ontológico de Tipo Matemático**, que pretende acercarse a esta idea usando LATEX® como editor de expresiones, y que permite usar un filtrado

Ontológico directamente en internet y en un repositorio de información.

El diseño del navegador web contextual para el campo de las matemáticas permitirá que los usuarios busquen con mayor facilidad la información requerida.

Para su desarrollo se deberán identificar los requerimientos de diseño de la herramienta de navegación, posteriormente un adecuado diseño y hacerse una prueba de contrastación para verificar su comportamiento.

- El diseño de la herramienta a partir del análisis de requerimientos de los usuarios brindará los elementos necesarios para la validación y puesta en marcha del software.
- Por un proceso de contrastación permitirán validar la funcionalidad, eficiencia y efectividad del navegador.

Entre las razones que justificaron investigar sobre este tema está la necesidad de mejorar contextualmente una consulta web de tipo matemático, optimizando su precisión, de forma que sean menos los resultados, pero con más significado, mejorando en los siguientes aspectos: Relevancia: porque la navegación contextualizada de tipo matemático genera un mejor resultado en la búsqueda de contenidos para el usuario; Eficiencia: porque implica el uso de filtrado para facilitar las búsquedas; Personalización: porque la búsqueda contextual puede emplear

la búsqueda que está almacenada en el Tesaurus (historial de navegación) como en la web.

Como puede apreciarse La relevancia de este trabajo radica en que se brinda un aporte hacia la navegación web contextual de tipo matemático, es decir, posibilita realizar una consulta a partir de ecuaciones, usando LATEX, Ontología y la necesidad del usuario, aspecto que hasta el momento comienza a ser nicho de investigación.

CAPITULO 1

1.1. Estado del Arte.

En este apartado se pretende mostrar un recorrido por las investigaciones, que han cimentado la base teórica abiertamente aceptada sobre la búsqueda contextual en internet, a partir de la cual se puedan conocer los avances en materia investigativa sobre el tema.

A continuación se muestra una tabla de investigaciones relacionadas con la búsqueda contextual de investigaciones de tipo matemático, que muestra autores, año de publicación, tema tratado, aportes a la investigación y limitaciones, la cual fue elaborada por el autor de la presente investigación a partir de la indagación de diferentes fuentes de información.

(Ver Tabla en la siguiente página).

TABLA 1. Investigaciones Relacionadas con la Búsqueda Contextual en Matemáticas Discriminada Cronológicamente.

TABLA DE INVESTIGACIONES RELACIONADAS CON LA BÚSQUEDA CONTEXTUAL EN MATEMÁTICAS DISCRIMINADA CRONOLÓGICAMENTE

AUTOR(ES)	AÑO DE PUBLICACIÓN	FUENTE	TÍTULO	TEMÁTICA	APORTES A LA INVESTIGACIÓN	LIMITACIONES
Natho, S. & Jeschke, S. et al.	2008	http://www.researchgate.net/profile/Sabina_Jeschke/publication/221573569_Information_extraction_from_mathematical_texts_by_means_of_natural_language_processing_techniques/links/0deec521be5299e5390000000.pdf	Natural Language processing methods for extracting Information from mathematical texts	Manejo de entidades textuales basadas en un esquema de clasificación lingüística para extracción de información a partir de un tesaurus y una herramienta que elabora un análisis de entrada, un análisis sintáctico, semántico y una integración de resultados en una base de conocimiento	Muestra como a partir de técnicas de procesamiento de lenguaje natural y el desarrollo de una herramienta inteligente, se puede obtener información de un tesaurus de tipo matemático	Las consultas se hacen solamente en un tesaurus y no en la red.
Natho, S. & Jeschke, S et al.	2008	eprints.physik.tu-berlin.de/159/1/PID505466.pdf	KEA A Mathematical knowledge Manager System combining Web 2.0 with Semantic Web Technologies	En este trabajo los autores presentan un sistema de Gestión de Conocimiento que sugiere la combinación de tecnologías Web 2 0 y Web Semántica. Este sistema semi-automático construye, organiza y almacena información en una base de conocimiento a partir de procesamiento de lenguaje natural PLN	Desarrollo de bases de conocimiento de información, es decir crea un tesaurus.	Las consultas se hacen solamente en un tesaurus y noen la red.
Wolska, Magdalena	2008	https://eudml.org/doc/220251	A language engineering Architecture for Procesing Informal Mathematical Discourse.	Se desarrolla una arquitectura para el procesamiento de texto matemático escrito sin estructura forma, tal como se encuentra en los textos y publicaciones matemáticas.	Se evidencia la complejidad de interpretar texto matemático y mucho más si este no tiene una estructura formal estandarizada en su publicación, tal como se encuentra en internet y como abordan este problema.	Aunque en este trabajo hay una mayor proximidad a la búsqueda a partir de símbolos, no hay pruebas en torno a la consulta directamente en la red.
Solovyev, V. & Zhitsov, N.	2011	http://dl.acm.org/citation.cfm?id=1988713	Logical Structure Analysis of Scientific Publications in Mathematics	Planteamiento de una aproximación y al análisis de la estructura de los artículos científicos matemáticos.	Desarrollo de la ontología Mocassin, empleada para extraer información de publicaciones académicas de tipo matemático	No identifica ecuaciones, solo texto de tipo matemático, todo se hace en una base de datos, no en la red.
Netzorova, O et al	2011	https://www.google.com.co/?gws_rd=ssl#	Bringing Math to LOD: A Semantic publishing platform prototype for scientific collections in mathematics.	Se desarrolla un prototipo de software para minería de artículos académicos matemáticos con el objetivo de obtener una representación de datos vinculados	Desarrollan un enfoque holístico para el análisis de documentos matemáticos así como una ontología de extracción, conversión y metadatos, así como la integración de algunos conjuntos de datos LOD (Link Open Data) existentes y búsqueda semántica.	Es de los trabajos más cercanos a la búsqueda semántica a partir de documentos matemáticos, aunque no se incluye la simbología en la consulta.
Yokoi, Keisuke et al.	2011	http://polibits.gelbukh.com/2011_43/43_11.pdf	Contextual Analysis of Mathematical Expressions for advanced Mathematical Search	Los autores trabajaron en una manera de como navegar en un documento publicado de tipo matemático a partir de computadoras. Dado que la información superficial de las expresiones es ambigua, estos también tienen en cuenta el texto adicional las contextualiza.	Extraen información a partir de técnicas de PNL, para nombres de variables o funciones que hacen parte de las expresiones matemáticas con diversos resultados. Emplean 2 técnicas de inteligencia artificial: contraste de patrones y máquina aprendente.	Toda la investigación se hizo para el idioma japonés.
Garcés, S. et al.	2012	https://www.google.com.co/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#	Modelo de búsqueda Web basado en información del Contexto del usuario y técnicas de filtrado colaborativo	Desarrollan un metabuscador Web que integra el filtrado colaborativo (basado en ítems) que recupera documentos de buscadores tradicionales como Google y Bing y los resultados se muestran listado en orden de relevancia, basado en la información del conteto del usuario y en la retroalimentacion colaborativa de la comunidad.	Este modelo constituye en un aporte para el área de recuperación de información, dado que muestra promisorios resultados en pruebas realizadas sobre colecciones cerradas y con usuarios.	Aunque la los resultados mostrados van orientado a investigaciones matemáticas y los resultados un nivel alto de acertividad, no hacen la consulta a partir de ecuaciones, sino plantean algoritmos de consulta, sin tener en cuenta las expresiones matemáticas

UNIVERSIDAD SIMÓN BOLÍVAR
BIBLIOTECA
INSTITUTO DE POSTGRADO

1.2. Marco Teórico Conceptual

1.2.1. Web Semántica

En la actualidad realizar una consulta en internet en donde la información no se encuentra organizada y clasificada por categorías de manera adecuada hace que los resultados esperados no sean los más óptimos. De hecho, desde los años 90 cuando se presenta oficialmente la World Wide Web que fue basada en documentos para ser leída por humanos y no comprendida por las computadoras no se tuvieron en cuenta los contextos de las búsquedas para que estas fueran mucho más precisas a la hora de utilizar un navegador cualquiera que este sea. (Berners-Lee Tim, 2001)

Por el problema antes descrito a finales de los años 90 aparece el concepto de Web Semántica, introducido por el Físico Británico Tim Berners-Lee (creador de WWW (Word Wide Web), de URIs (*Uniform Resource Identifier*), de la HTTP (*Hypertext Transfer Protocol*) y HTML (*Hypertext Markup Language*)) definiéndola de la siguiente manera: "*El primer paso es colocar los datos en la Web de manera que las máquinas puedan entenderlo "naturalmente" o convertirlos de esa forma. Esto crea lo que yo llamo una Web Semántica: Una red de datos que pueden ser*

procesados directamente o indirectamente por máquinas" (Tim B.-L. , 2000).

La Web semántica busca mejorar la relación entre las computadoras y los seres humanos, ofreciendo a las máquinas mayor autonomía e inteligencia y a las personas herramientas que le ayude a la construcción del conocimiento a partir de información mucho más organizada. (Cueva, 2008).

Tim Berners-Lee comenta "*no se está hablando de una inteligencia artificial mágica que permita a los computadores entender palabras de los usuarios, sino únicamente de la habilidad de la máquina de resolver problemas bien definidos, a través de operaciones bien definidas que se llevarán a cabo sobre datos bien definidos*" (Tim B.-L. , 2000).

Actualmente el común de las páginas web están diseñadas en lenguaje HTML careciendo de metadatos semánticos lo que permitiría que cualquier usuario pueda hacer búsquedas de manera eficiente en internet y así encontrar más eficientemente la información requerida debido a que estará mejor definida. Los metadatos deberán describir el sentido, el contenido, el significado y la relación entre estos, mejorando sustancialmente la búsqueda de información debido a la estandarización de la infraestructura, permitiendo organizar,

transferir, compartir, procesar cualquier tipo de información de manera mucho más simple.

El objetivo de la web semántica es mejorar la interacción entre computadoras, y reduciendo intervención de seres humanos, para ello deberá emplear agentes inteligentes que serán los encargados de extraer información de muchas fuentes de forma que la consulta arroje los resultados esperados, combinando información, razonando con los datos encontrados y finalmente haciendo deducciones lógicas con lo encontrado automáticamente (Cueva, 2008).

1.2.2. Conceptos Relacionados con Web Semántica.

Por el gran número de publicaciones que hay en internet, la consulta se ha convertido en un proceso complicado, de hecho, basta con aplicar un comando en Google del web master tools para buscar el número total de indexaciones que tiene una página web que dejan de lado información importante, en especial aquellas que están hechas de manera dinámica.

Con base en esto, se manejan 2 conceptos a saber: **Web Superficial**, que se refiere al listado de páginas que existen en red que son indexables y **Web Profunda** que es el conjunto de páginas que no se encuentran definidas en un servidor, sino que son el resultado de una aplicación diseñada en PHP, JSP, etc., debido a que estas son hechas en tiempo real y que dependen de las características de solicitud de información de un usuario.

- **Arquitectura de una Web Semántica.**

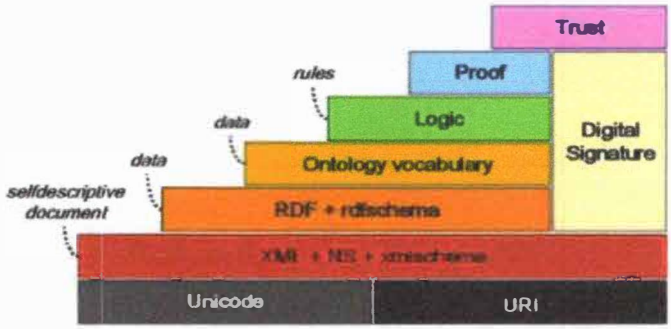


Gráfico 1. **Modelo de Arquitectura de la Web Semántica según World Wide Web Consortium**

http://www.w3.org/2001/09/21-orf/hagino-sw/swlevels.gif

La arquitectura de la red semántica ha sido organizada por niveles, donde se establece un esquema de abstracción y relaciones entre los diferentes niveles. Este es el modelo inicial con el que fue concebida y actualmente los avances llegan hasta el nivel de lógica (ver gráfico).

A continuación se muestra una tabla que describe los siete niveles de la Web Semántica:

Los 7 Niveles de la Web Semántica

NIVELES	DESCRIPCIÓN	DESCRIPCIÓN DE SUS COMPONENTES
1. NIVEL DE RECURSOS	Identifica el nivel de importancia de los recursos que se encuentran distribuidos en la red. En este nivel se destacan 2 aspectos: **Unicode** y **URI**	**UNICODE:** Es un estándar cuyo objetivo es proporcionar que cualquier texto en cualquier idioma sea codificado sin ningún problema.
		URI(*Uniform Resourse Identifier*): Se encarga de identificar las diferentes páginas web sin problemas. Estas cadenas permiten acceder a cualquier recurso de la web.
2. NIVEL SINTÁCTICO	En esta etapa se solciona el problema de como definir distintos lenguajes de etiquetados para añadir contenido semántico a las páginas web. Sus componentes son **XML+NS+XMLSCHEMA**	**XML (*Extensible Markup Language*):** Esta es la capa más técnica de la Web Semántica. En este nivel se encuentran agrupadas las diferentes técnicas que posibiliten la comunicación entre agentes. XML proporciona un formato comín para el intercambio de documentos
		NS (*Namespace*): Como su nombre lo indica, proporciona un método para evaluar elementos y atributos de nombres usados en documentos XML asociándolos con espacios de nombre identificados por referencias URIs.
		XMLSCHEMA: XML Schema es un lenguaje que describe la estructura y restringe el contenido de documentos XML.
3. NIVEL DE DESCRIPCIÓN DE RECURSOS	Se define el lenguaje universal con el cual se expresan diferentes ideas de la Web Semantica. Sus componentes son: **RDF+RDFSCHEMA**	**RDF(*Resource Description Framework*):** Sugerido por la W3C (World Wide Web Consprtium) diseñada para estandarizar la definición y el uso de las descropciones de metadatos y representación datos simples de los recursos Web: usuarios dispositivos, páginas, etc. Este lenguaje de etiquetado está basado en XML y ofrece estructuras formales gracias al uso de URI's, permitiendo la codificación, intercambio y procesamiento de meta-datos normalizados.
		RDFSCHEMA: Es una extensión de RDF que aporta un lenguaje con mayor capacidad para representar relaciones semánticas complejas
4. NIVEL DE ONTOLOGÍAS	Una ontología es una especificación formal de un dominio del conocimiento que, en su expresión más simple, se identifica con una taxonomía. Una taxonomía consiste en una jerarquía de conceptos y sus relaciones del tipo clase-subclase. Una ontología formaliza la relación declase, añade otras relaciones y especifica propiedades para individuos y clases. Ontology-vocabulary se refiere a una ontología concreta sobre un dominio concreto del conocimiento. El conocimiento no solo debe ser legible por una máquina , sino que debe ser de común acuerdo y reutilizable. El Lenguaje de la Ontología es el ***OWL (Ontology Web Language)*** y consta de 3 sublenguajes que son usados depende las necesidades **OWL lite, OWL dl** y **OWL full**.2	**OWL lite:** Este sublenguaje clasifica los conceptos jerárquicamente y con restricciones sencillas. Esta versión fue creada para representar las necesidades básicas de la ontología.
		OWL dl: Su objetivo es hacer la descripción de conceptos del dominio y la semántica que establece una equivalencia entre las fórmulas de lógicas de descripción y expresiones en lógica de predicados de primer orden.
		OWL full: Permite la misma expresividad que OWL dl y tiene la misma capacidad sintáctica de RDF, sin embargo agrega una mejora en el procesamiento debido a que preserva compatibilidad con RDF+schema. Por ejemplo una clase con OWL full puede tratar a una clase como un conjunto de individuos, pero también puede tratarla como una entidad por derecho propio, además este lenguaje puede usar bucles infinitos en el razonamiento.
5. NIVEL DE LÓGICAS	En este nivel se tiene en cuenta las reglas de inferencia lógica. Una ontología puede expresar una regla tal como: "Si un código de una ciudad, está asociado a un código de estado, y si una dirección es el código de la ciudad, entonces esa dirección tiene el código de estado asociado". De esta manera la Universidad Simón Bolivar al estar en Barranquilla, por ende está ubicada en Colombia y esto da muestra que se encuentra definida bajo los estándares de la W3C. La máquina no va a entender la información, pero si puede hacer que la información sea mucho más clara para un usuario. El nivel de Lógica tiene por objetivo flexibilizar la arquitectura web para la realización de consultas y poder inferir conocimientos usando las ontologías del nivel anterior.	
6. NIVEL DE PRUEBAS	Depende del nivel anterior debido a que debe realizar pruebas usando el lenguaje unificador de la Web Semántica. Este lenguaje permite que las inferencias lógicas que se hagan mediante el uso de las reglas de inferencia del nivel anterior.	
7. NIVEL DE CONFIANZA + FIRMA DIGITAL	Esta es una etapa de verificación de las etapas anteriores, es decir, hasta que no se haya verificado el origen de la información, los agentes inteligentes estarán escépticos de lo encontrado en la web semántica, es decir, debe verificar la firma digital para tal efecto, (Digital Signature) de una fuente confiable de información y la firma digital es un bloque de datos encriptados que son usados por las computadoras que verifican que la información obtenida, sea de una fuente específica confiable.	

ARQUITECTURA DE LA WEB SEMÁNTICA

Tabla 2. Los 7 Niveles de la Web Semántica (Berners_lee Tim, 2002).

1.2.3. Elementos de la Web Semántica

Los elementos básicos que componen la web semántica son: XML, RDF, PICS, ONTOLOGÍAS y AGENTES. (Ver Tabla 3.)

ELEMENTOS DE LA WEB SEMÁNTICA	
XML	(Extensive Markup Language), lenguaje que facilita la codificación para la distribución de documentación compleja por Internet, generando una estructura mas semántica en la Web. Permite que los diseñadores de páginas web, usen sus propias etiquetas de demarcación (markup-tags). Esas etiquetas son elegidas de forma tal que denotan la propia semántica del dominio tratado, aparte de encargarse de la posición específica y las características propias de la información que contiene la página web. XML es un subconjunto de SGML (Acrónimo en inglés Standard Generalized Markup Language, es decir, Estandar de Lenguaje de Marcado Generalizado) y define un formato de texto diseñado para la transmisión de datos estructurados. Al ser un subconjunto de SGML conserva las características de validación, estructurado y extensibilidad, debido a que es un metalenguaje que describir lenguajes de marcas, tanto la definición de etiquetas como la relación estructural que existen entre ellas.
RDF	Resource Description Framework). Describe los recursos de la red, es decir, páginas, personas, dispositivos, objetos, imágenes, etc., ajustando las condiciones necesarias para que lo consultado haga parte de las características definidas en el sistema, de forma que no genere ambigüedades, por el uso de los URI, que permite codificación, intercambio y procesamiento de los metadatos que han sido estandarizados. Este modelo se basa en la idea de convertir las declaraciones de los recursos en expresiones con la forma sujeto-predicado-objeto (conocidas en términos RDF como tripletes). El sujeto es el recurso, es decir aquello que se está describiendo. El predicado es la propiedad o relación que se desea establecer acerca del recurso. Por último, el objeto es el valor de la propiedad o el otro recurso con el que se establece la relación. La combinación de RDF con otras herramientas como RDF Schema y OWL permite añadir significado a las páginas, y es una de las tecnologías esenciales de la Web semántica. (W3C: World Wide Web Consortium, 2004) Una de las implementaciones más populares de RDF es la que se conoce como las páginas RSS (Really Simply Sindication) estas páginas están fundamentadas en formato XML para compartir contenido en la Web. Gráfico 2. Icono de las páginas RSS http://www.dewartrocco.es/2011/10/que-es-un-feed-rss-y-como-crear-el-feed.html
PICS	Los PICS (Platform for the Internet Content Selection), Permite ver características específicas de archivos dependiendo de la comunidad en la que se encuentre el usuario. Es una infraestructura que asocia las etiquetas con los contenidos de Internet, tales como privacidad, licencia, acceso restringido, etc. PICS es una plataforma en donde se han

	construido otros servicios de clasificación que no sólo define una manera de construir etiquetas sino que es un mecanismo para realizar las valoraciones. Este mecanismo está formado, al menos por lo siguiente: Las etiquetas, que son los metadatos, indican la valoración de un documento. Los servicios de valoración, es decir, las organizaciones, grupos o personas que realizan una valoración. Los perfiles, que son las reglas que da el usuario para definir el filtro y evitar recibir documentos no deseados. (Web Semántica: Agentes Inteligentes, 2004) Para que el filtrado de documentos no deseados se lleve a cabo, también es necesario un software cliente y otro servidor que tengan implementado el sistema de valoración. Estas funciones se pueden realizar por separado, lo cual permite que por un lado los desarrolladores de software puedan realizar una aplicación informática sin suministrar un sistema de valoración mientras que por otro una organización puede crear sistemas de valoración sin tener que desarrollar el software. (Web Semántica: Agentes Inteligentes, 2004)
ONTOLOGÍAS	"las ontologías son colecciones de enunciados redactados en un lenguaje, como el RDF, que define las relaciones entre conceptos y especifica reglas lógicas para razonar con ellos. Las computadoras "comprenderán" el significado de los datos semánticos de una página de la red siguiendo vínculos con ontologías especificadas". (Carolina García Cataño, 2002)
AGENTES	Desde hace algunos años, los agentes son considerados una herramienta de uso creciente, esto se debe a la evolución de la informática y las telecomunicaciones o lo que se conoce como Telemática. (Uribe Tirado, 2004) Según Maes, P. "*los agentes son sistemas computacionales que habitan en entornos dinámicos complejos, perciben y actúan de forma autónoma en ese entorno, realizando un conjunto de tareas y cumpliendo objetivos para los cuales fueron diseñados*". En la web semántica cumplirán la función de buscar servicios, para ello, la semántica hará que los agentes describan unos a otros la función exacta que realizan, y qué datos deberán recibir. El comportamiento de los agentes inteligentes muestra que son programas avanzados que funcionan para los usuarios en consultas de manera eficiente.

Tabla 3. Varela Arregocés E. Elementos de la Web Semántica.

Tabla realizada por el Investigador.

1.2.4. Servicios de la Web Semántica

Con los Servicios Web Semánticos estamos ante una nueva aproximación que consiste en la integración de la Web Semántica y los Servicios Web como tecnologías complementarias.

Gráfico 3. Servicios de la Web Semántica
Tomado de: La Web Semántica. Santiago Márquez Solís 2007.

Los Servicios Web Semánticos proporcionan una nueva infraestructura para los Servicios Web al añadir la posibilidad de definir la semántica de los mismos.

De este modo los problemas que se han visto cara al descubrimiento de los Servicios Web y de las capacidades que presentan podrán resolverse de manera automática o semi-automática.

Algunas ventajas que se obtendrán con el uso de esta tecnología son las siguientes:

- Buscadores basados en Web Semántica para Servicios Web: Por ejemplo, podremos realizar búsquedas basadas en restricciones complejas del tipo "encontrar el hotel más bonito cercano a la playa".

- Se tendrá la posibilidad de proporcionar un Servicio Web que haga deducciones lógicas por mí, por ejemplo comprobar la validez de la descripción de un Servicio Web. (Márquez Solís, 2011).

1.2.5. Aplicaciones Técnicas de la Web Semántica

Entre los campos de aplicación donde las nuevas ideas de la web semántica pueden tener utilidad podemos citar: comercio electrónico, gestión del conocimiento corporativo, enseñanza, librerías digitales, turismo, patrimonio cultural, entre otras, se destacan en especial que son la búsqueda de información en la web y el procesamiento del lenguaje natural que se describen a continuación.

- **Búsqueda de información en la web**: Por ejemplo, una investigación titulada: "Sistema de Navegación Colaborativa basada en mashup semántico con API's abiertos" (Collaborative browsing system based on semantic mashup with open API's). En este trabajo el autor propone una aplicación web semántica (mashup) que realiza navegación en colaboración (co-navegación) usando "Open API'S", se comparte información con otros agentes

intercambiando marcadores, selecciona el más confiable y a su vez integra información heterogénea encontrada. (Jung, 2012)

- **Procesamiento del lenguaje natural**: En este caso se menciona una herramienta llamada CALAIS. El "OpenCalais Web Service", crea automáticamente ricos metadatos semánticos de acuerdo al contenido que se envíe, en menos de un segundo, usando Procesamiento de lenguaje natural (en Inglés), aprendizaje automático y otras metodologías. CALAIS analiza el documento y encuentra las respectivas entidades dentro de ella y devuelve los aspectos ocultos del contexto. (Buitelaar, 2009) y (Reuters, 2008)

UNIVERSIDAD SIMON BOLIVAR
BIBLIOTECA
INSTITUTO DE POSTGRADO

1.2.6. Ventajas y Desventajas de la Web Semántica

VENTAJAS	DESVENTAJAS
• La principal ventaja que aporta la Web Semántica es que incorpora contenido semántico a los documentos que se colocan en Internet, permitiendo organizar por conceptos la información que ofrecemos, garantizando búsquedas sobre esta por significado y no por contenido textual (Leguízamo León, 2007). • Permite a las computadoras la gestión, de conocimiento , hasta el reservada a las personas (inteligencia artificial) • Ahorro de tiempo en el procesado de datos (tiempo de búsqueda, gestión de la información, etc): la mayor parte de las tareas relativas al procesado de datos podrán realizarse por parte de componentes software automatizados sin requerir, en muchos casos, intervención humana. • Mejora de comunicación de los servicios web: la comunicación entre distintos componentes y servicios, sobre todo en aquellos casos en los que los componentes no han sido diseñados para trabajar conjuntamente, ha sido siempre fuente de problemas en cuanto a la interoperabilidad debido principalmente a la ambigüedad del lenguaje. El uso de ontologías compartidas (y, posiblemente, mapeadas con otras	• La poca efectividad de los sistemas basados en reglas de inferencia, debido a que es uno de los requisitos que exige la web semántica. • Si se pudiera resolver el problema de las reglas de inferencia, la implantación de una infraestructura de redes de confianza, es decir que sea factible el intercambio de información de manera automática a costos razonables y que el nivel de complejidad de la creación de esta sea manejable. • El cambio de paradigma de los usuarios y de los creadores de páginas web. • El costo de transición de la Web actual a la Web Semántica. • El diseño y creación de Ontologías. (Márquez Solís, 2011)

ontologías) propuesto por la Web Semántica, solventa el problema en la comunicación entre servicios Web gracias a que esta comunicación se produce gracias a conceptos pertenecientes a la ontología.	

Tabla 4. Varela Arregocés E. Ventajas y Desventajas de la Web Semántica. Tabla elaborada por el investigador

1.2.7. Ontologías

Al principio de los 90's con el desarrollo de internet, las Ontologías se convirtieron en objeto de estudio para muchas áreas del conocimiento: Inteligencia artificial, Ingeniería del Conocimiento, Procesamiento del lenguaje natural, representación del Conocimiento, recuperación de información, gestión de conocimiento y servicios web, entre otros aspectos. Esta herramienta se populariza debido a que puede alcanzar una común y compartida comprensión de algún dominio en particular y que puede ser comunicado entre usuarios y máquinas, permitiendo así que muchos procesos puedan ser automatizados.

La creación de una ontología no es un proceso al azar, se deben seguir una serie de pasos que llevan un orden metodológico y de hecho no existe una sola manera de realizarlo. Todas las ontologías no están creadas, por ello deben hacerse para que puedan ser de uso común.

Basado en lo anterior, surgen algunas preguntas ¿A cuáles herramientas de desarrollo se puede acudir para implementarla? ¿Con qué lenguaje pueden ser marcadas? Además, en qué orden se debe aplicar cada elemento (metodología, Herramienta y Lenguaje) y que relación guardan unos con otros.

- **Definición de Ontología:**

Ontología es una disciplina proveniente de la filosofía, de la rama de la metafísica que estudia todo lo existente y de cómo cada uno de los entes, se relacionan (Hofweber, 2004). Desde el concepto de la Informática, se refiere a diferentes teorías que especifican un léxico relacionado con cierto dominio. Este léxico define un listado de entidades, clases, propiedades, predicados lógicos, funciones y las relaciones entre cada uno de estos aspectos. Algunas definiciones desde el ámbito de la informática son las siguientes (ver Tabla):

ALGUNOS CONCEPTOS DE ONTOLOGÍA EN INFORMÁTICA

AUTOR(ES)	CONCEPTO
Robert Neches y otros autores	"Una ontología define las condiciones básicas y relaciones que comprenden el vocabulario de un área del tema así como las reglas para combinar condiciones y las relaciones para definir extensiones del vocabulario"; según la definición, los autores indican los pasos para hacer una ontología, pero su explicación es bastante somera, pues solo se indican las condiciones básicas y las relaciones entre las condiciones, permite que se identifiquen las normas para las combinaciones de estas, y brinda las conceptos de estas condiciones y relaciones. Se puede destacar que los autores hacen mención de las condiciones que se hacen explícitas en la ontología, sino que se incluye el conocimiento que puede ser inferido. (Neches 1991)
Thomas R. Gruber	Una especificación explícita de una conceptualización, es decir, que proporciona una estructura y contenidos de forma explícita que codifica las reglas implícitas de una parte de la realidad; estas declaraciones explícitas son independientes del fin y del dominio de la aplicación en el que se usarán o reutilizarán sus definiciones". (Gruber, 1992).
World Wide Web Consortium W3C	Una ontología define los términos a utilizar para describir y representar un área de conocimiento. Las ontologías son utilizadas por las personas, las bases de datos, y las aplicaciones que necesitan compartir un dominio de información (un dominio es simplemente un área de temática específica o un área de conocimiento, tales como medicina, fabricación de herramientas, bienes inmuebles, reparación automovilística, gestión financiera, entre otras). Las ontologías incluyen definiciones de conceptos básicos del dominio, y las relaciones entre ellos, que son útiles para los computadores. Codifican el conocimiento de un dominio y también el conocimiento que extiende los dominios. En este sentido, hacen el conocimiento reutilizable

Tabla 5. Varela Arregocés E. Algunos Conceptos de Ontología en Informática

Tabla realizada por el Investigador.

Aunque se mostraron algunas definiciones formales de ontología, todos los seres humanos las tenemos en nuestro pensamiento, con estas se representa y entiende todo el mundo que lo rodea. No necesariamente estas tienen que estar totalmente detalladas, ni tampoco mostrar el comportamiento matemático relacionado con estas, por ejemplo, si se menciona un avión, de forma inmediata se piensa en un medio de transporte

aéreo que lleva gran cantidad de pasajeros que viajan de una ciudad a otra, y para esto, no se hizo necesario formalizar la ontología involucrada, debido a que esta es una información cotidiana y que muchas personas comparten.

De igual manera ocurre cuando se habla del dominio familiar, en donde se sabe que existe una organización arbórea y jerárquica de la organización, en donde un hijo, tiene un padre y una madre y esto a su vez con cada uno de ellos, por otro lado, los hijos (o hijas), tienen hermanos (o hermanas), además, un hijo (o hija) puede tener una sola madre biológica y un padre biológico, son conceptos que todo el mundo conoce y maneja y no se hace necesario definirlos. En cambio, cuando se habla de terminología desconocida, o si se menciona que sea procesada por una computadora, se deben emplear ontologías para que las máquinas puedan procesar la información. Las máquinas en términos generales carecen de ellas, por eso deben ser creadas, para que puedan "entender la información" y así procesarla, para tal efecto hay diferentes métodos y metodologías para su creación y desarrollo.

- **Terminología Relacionada con las Ontologías**

Una Ontología expresa el conocimiento de los dominios a través de los siguientes componentes (Velázquez Muñoz, 2008) y (Contreras, 2007): (Ver Tabla)

TERMINOLOGÍA RELACIONADA CON LAS ONTOLOGÍAS	**Clases:** son los conceptos de un dominio que se desea representar. Es el aspecto más relevante a la hora de hacer estructuración del conocimiento, debido a que depende de los objetivos de la Ontología. Normalmente una clase puede tener subclases que pueden representar conceptos que son más específicos que la misma clase.
	Clases Abstractas: Clase que no permite que existan instancias de ella. Se usa para agrupar conceptos, introducir cierto orden en la jerarquía, pero suelen ser demasiado generales para admitir instancias.
	Instancias: son los objetos de una clase. El conjunto de Instancias constituyen la base de datos de conocimientos del modelo.
	Instancias Indirectas: Cuando una clase es instancia indirecta de otra, quiere decir que es instancia de alguna de sus clases derivadas. En contraposición a instancia directa, donde no existen clases intermedias.
	Relaciones: Representan la conexión entre clases Interacción o enlace entre los conceptos o clases del dominio que se modela. Algunas relaciones semánticas básicas son: subclase de, parte de, parte exhaustiva de, conectado a, es un, etc. Suelen configurar la taxonomía del dominio. Las relaciones más simples se modelan mediante una propiedad de una clase cuyo valor es una instancia de otro concepto.
	Funciones: Identifican elementos a partir de algún cálculo que hacen parte de una ontología. Por ejemplo: *identificar _fecha, clasificar _elemento*, etc.
	Axiomas: Se refiere a las restricciones que se aplican a ciertos elementos de una ontología, para especificar claramente el dominio del conocimiento.
	Anotación: Es el proceso de relleno de instancias a partir de texto libre. Existen dos maneras de anotar texto: la más usual es la inclusión de etiquetas semánticas dentro del texto que se está procesando. Esto implica que el formato del texto debe ser editable y procesable. En el caso de no disponer de este formato, la segunda manera consiste en rellenar las instancias directamente en el modelo, dejando el texto original sin modificar.
	Herencia: Es una propiedad de la relación "es_un" que permite que las clases relacionadas (heredadas) cuenten con los atributos de la clase con la cual se relacionan (clase padre).
	Herencia Múltiple: Se da cuando una clase dada hereda los atributos de dos clases padre con las que establece dos relaciones del tipo "es_un".
	Derivación: Organización de las clases de la ontología en un árbol de jerarquía mediante sucesivas relaciones "es_un" con la propiedad de la herencia. Esta organización permite el encadenamiento sucesivo de herencias desde las clases de nivel superior a las clases situadas en niveles inferiores, llamadas clases derivadas.

Tabla 6. Varela Arregocés E. Tabla de Terminología Relacionada con las Ontologías

Tabla realizada por el investigador.

- **Aspectos a tener en cuenta para el diseño de una Ontología**

Cuando se pretende usar una Ontología para la representación de un dominio, deben tomarse algunas decisiones para el diseño (Velázquez Muñoz, 2008):

- Claridad: Una ontología debe comunicar eficazmente que significan los términos definidos, estos deben totalmente objetivos y deben expresarse en el lenguaje más natural posible.
- Coherencia: Una ontología es incoherente si una frase que se deduce a partir de axiomas, contradice una definición
- Extensibilidad: Una condición es la definición de nuevas expresiones a partir del vocabulario existente en las ontologías de manera que no se requiera la descripción de nuevos términos.

- **Áreas de Aplicación de Ontologías**

Las posibles aplicaciones en el empleo de Ontologías son entre otras (Velázquez Muñoz, 2008): Repositorios para la organización y comparación del conocimiento, Herramientas para la adquisición de Información, aplicaciones destinadas a favorecer la interoperabilidad entre sistemas heterogéneos, Motores de Búsqueda semánticos, herramientas de referencia en

la construcción de sistemas de bases de conocimiento que aportan consistencia y fiabilidad a la hora de recuperar los datos, aplicaciones destinadas al trabajo colaborativo al funcionar como soporte común de conocimiento entre organizaciones, comunidades científicas, entre otras.

- **Reutilización de Ontologías**

De los aspectos importantes al representar información usando ontologías, es la reutilización. Este aspecto es una ventaja porque reduce costos en el desarrollo, se logra la interoperabilidad entre otros sistemas, etc. (Velázquez Muñoz, 2008).

Entre los repositorios de ontologías existentes más importantes se destacan: SOUPA (Ontología Estándar para Aplicaciones Ubicuas Dominantes, Acrónimo en Inglés de *Standard Ontology for Ubiquitous and Pervasive Applications*) (Chen, Finin, & Joshi, 2005), cuyo objetivo es orientar a los desarrolladores que no tienen experiencia en la representación del conocimiento a construir rápidamente aplicaciones basadas en ontologías. Otro modelo de repositorio es CONON (Ontología Basada en Modelado de Contexto y Razonamiento usando OWL. Acrónimo en inglés de *Ontology Based Context Modeling and*

Reasoning using OWL), que proporciona una ontología orientada a aplicaciones sensibles al contexto (Wang, Gu, Zhang, & Pung, Marzo 14 de 2004); FIPA (*Foundation for Intelligent Physical Agents) Device Ontology* utilizada para expresar las capacidades de diferentes dispositivos en sistemas de computación Ubicua (Foundation for Intelligent Pshysical Agents, 2001); GUMO (El usuario común del modelo ontológico, Acrónimo en inglés *The General User Model Ontology)* (Heckmann, Schwartz, Brandherm, & Kröner, 24-29 de Julio 2005) Ontología para el modelado de conceptos generales, entre otras.

- **Metodologías y Métodos para el diseño de una ontología**

Para diseñar una ontología se hace necesario tener una metodología que indique los pasos para representar la información. Cada metodología tiene sus propias características, la gran mayoría comparten las siguientes propiedades (Velázquez Muñoz, 2008):

- Ciclo de Vida: Es el periodo que transcurre desde la concepción, elaboración y mantenimiento de la ontología.
- Herramientas de Apoyo: Las metodologías se apoyan en herramientas ontológicas que faciliten la creación de Ontologías.

- Nivel de Modelado: Da una idea sobre cómo se realiza la extracción de los datos
- Nivel de Abstracción: Da una idea del nivel de relación existente entre la información concreta y la forma de extraerla. Cuanto mayor sea el nivel de abstracción, más fácil será para las aplicaciones extraer los datos.
- Uso de Ontologías Base: habitualmente se reutilizan términos o características de otras que ya han sido creadas.

1.2.8. Contexto Definitorio (CD)

Una de las actividades que ha tenido éxito es el reconocimiento de la extracción de términos textuales de manera automática (Cabré, Estopà, & Vivaldi, 2001). Sin embargo en lo relacionado con la extracción de conceptos implica una actividad bastante compleja debido a que manejan distintas posturas acerca de los límites y características de una definición.

Por lo anterior, el análisis lingüístico de una definicion no es un asunto trivial. Una manera de destacar la riqueza de relaciones que puede mantener una definicion respecto a un termino, es considerar que ambos se configuran en una estructura llamada contexto definitorio (CD), entendido este como un fragmento textual donde se introduce un término y su

UNIVERSIDAD SIMON BOLIVAR
BIBLIOTECA
INSTITUTO DE POSTGRADO

correspondiente definición (Alarcón & Sierra, 2003). Este trabajo se enfoca en la presencia de patrones lingüísticos y no-lingüísticos constitutivos de CDs,asi como en el uso de dichos patrones para identificar y etiquetar de manera automática términos y definiciones en textos especializados.

- **Elementos de un Contexto Definitorio (CD)**

Un Contexto Definitorio (CD) presenta estructuras diversas, pero en esencia cuenta con elementos tales como: término, definicion, predicación verbal definitoria, marcadores reformulativos definitorios, marcadores tipográficos definitorios, patrones pragmáticos y relaciones de correferencias y anaforas que se establecen entre el termino definido y estas unidades enlistadas. La identificación de dichos componentes permite distinguir patrones de comportamiento sintáctico y semántico que facilitan la obtención de datos para la búsqueda y clasificación de las definiciones.

Los dos componentes mínimos y basicos de un CD son el término y la definición del mismo. Se ha observado que ambos elementos establecen ligas a partir de predicaciones verbales definitorias (PVD) del tipo X se define como Y, X comprende Y,

X es Y, entre otras. Las PVDs mantienen una estructura reconocible en patrones determinados, llamados patrones definitorios (PD), que permiten establecer un catalogo de verbos asociados a tipos de definiciones especificas. La búsqueda y establecimiento de estos patrones verbales ha sido fundamental en el análisis de CDs.

Si bien estos tres elementos configuran una estructura canónica de un CD, no siempre ocurre asi, ya que la mayoria de las veces los CDs presentan otros PDs que también ligan a términos y definiciones, los cuales son (Gerardo Sierra, 2006):

- **Marcadores reformulativos definitorios** (MRDs): estructuras sintácticas relacionadas con un proceso metalingüístico. Estas construcciones ayudan a reinterpretar o retomar algún elemento discursivo para presentarlo de otra forma, p. e.: es decir, por ejemplo, esto es, etc.

- **Marcadores tipográficos definitorios** (MTDs): cualquier signo de puntuación o marca tipográfica que tenga como función, por un lado, ligar a un término con su definición, sustituyendo o complementando la función de la PVD (viñetas, paréntesis, guiones, comillas, etc.); y por otro lado, aquellas marcas que recalquen la presencia

tanto de un término como de su definición (negritas, subrayado, cursivas,comillas, etc.).

- **Patrones pragmáticos** (PPs): proporcionan información sobre el uso de los términos. Se consideran tres variantes: (a) autoría, cuando es un autor o una institución quien define el término; (b) patrones pragmáticos temporales, en el caso en que se proporcione alguna fecha en la que se definió o se introdujo el término; y (c) patrones instruccionales, referidos a estructuras sintácticas que dan matices diferentes para introducir la definición: desde el punto de vista, de manera general, desde la perspectiva, científicamente, etc.

Finalmente se plantea la presencia de correferencias y anaforas que, aunque no son componentes de un CD, implican relaciones referenciales a partir de las cuales se puede deducir si un termino esta ligado a otros terminos o entidades linguisticas alejadas del CD. Correferencias y anaforas tienen un contenido conceptual valioso para determinar el sentido del termino definido. (Gerardo Sierra, 2006)

Gráfico 4. Ejemplo de Un Contexto Definitorio (CD) Tomado desde: (Gerardo Sierra, 2006)

- **Contexto de una Búsqueda Web**

La personalización de las consultas web alivia la sobrecarga de información mediante procesos adaptación de la información presentada, basada en las necesidades del usuario.

Un aspecto importante para el adecuado acceso a la información personalizada es el contexto del usuario. Un sistema que no sabe qué tipo de información busca y para que propósito la consulta, solamente ofrecerá respuestas generales.

En los últimos años la búsqueda personalizada ha despertado el interés de la comunidad científica, como forma de reducción de la ambigüedad en los resultados de consulta web y en la obtención de precisión en las páginas obtenidas de

mayor interés para un usuario en particular; muy a pesar de esto, la interacción de los usuarios con los motores de consulta web que se caracterizan por su generalidad en los resultados, prácticamente no es tenida en cuenta, por ello es que se presentan inconvenientes de filtrado, ordenamiento y manejo del contexto. (Sara Garcés, 2012)

Los problemas más frecuentes que se presentan en las consultas web están: el volumen de información, el spam en los contenidos y en los vínculos, los lenguajes en que fueron diseñadas las páginas, la recuperación de información multimedia, poca calidad de interfaz de usuario y la pobreza en la visualización de los resultados de consulta (R. Baeza-Yates, 2006).

Respecto al contexto en los motores de búsqueda, este no es capturado en el tiempo de indexación, ni tampoco es usado en el tiempo de recuperación, lo anterior hace que el resultado de la consulta sea inexacta, como consecuencia de esto el usuario termina recibiendo información mixta, invirtiendo más tiempo en la lectura de las páginas recuperadas, ocasionando abandono de la búsqueda, sin obtener el resultado deseado. (Melucci, 2005).

Es importante recordar que el análisis del usuario en su contexto requiere la utilización de información que describa

sus acciones. Los datos de los que se dispone dada la tecnología subyacente a los motores de búsqueda corresponden a los logs del motor de búsqueda. En este punto puede ser considerada información adicional que incluye aquella que es generada por el sitio Web asociado al buscador y los datos que son derivados a partir de la información recogida en el log. El log del motor de búsqueda contiene el registro de las acciones del usuario mientras usa un buscador, este registro incluye la dirección IP del usuario, el momento en que se hace la consulta, la cadena de búsqueda que incluye la consulta, el número de resultados que son devueltos y el tiempo de respuesta del buscador. Pero extraer perfiles de usuario integrando estos datos con información de dominio es un reto hoy en día. (Peláez, 2007)

Existen muchos enfoques en los que este problema ha sido abordado entre estos están: cálculo de la relevancia de usuarios, análisis de los documentos vistos o editados por el usuario, mejora de la tecnología empleada en la búsqueda, retroalimentación implícita y explícita, categorización de usuarios y de consultas entre otros. (Peláez, 2007).

- **Análisis Contextual de Expresiones Matemáticas.**

El análisis contextual de una ecuación o de una expresión matemática es un tema bastante complejo de abordar debido a que esta por sí sola no tiene ningún sentido, si no es acompañado de una explicación que contextualice dichos símbolos. De igual manera ocurre en una consulta web, en donde al escribir una ecuación en un navegador, esta arroja como resultado un listado de direcciones que no orientan mucho al proceso de consulta, muy a pesar de que las matemáticas son consideradas como un lenguaje universal, empleado por la comunidad científica, no solo para generar cálculos sino expresar en forma de un lenguaje sintetizado dicho conocimiento.

Son pocos los trabajos desarrollados en torno a este tema, porque habitualmente al momento de consultar en la web o la expresión matemática es obviada o simplemente se aborda desde el texto que acompaña a la ecuación o expresión o quizá el autor de esta o en su defecto, esta es tratada como imagen, usando técnicas tales como machine-learning, identificación de patrones, entre otros. (Keisuke Yokoi, 2011)

Si un buscador abordara la interpretación de una ecuación sin aspectos que orienten el contexto de la búsqueda, no solo

se estaría enfrentando a una máquina de Touring en su estado más puro, sino que los resultados serían extremadamente ambiguos, mixtos, sin sentido, etc., de hecho normalmente una ecuación no puede ser escrita en el espacio de consulta de los buscadores, habituales, entre otras cosas, porque el teclado de los equipos de cómputo, fijos o móviles no cuenta con toda la simbología para ello.

La revisión bibliográfica para este apartado arrojó muy pocos resultados, porque lo encontrado muestra a lo más investigaciones que tienen en cuenta 2 aspectos uno el texto que acompaña a la expresión matemática como tal. Sin embargo, los documentos revisados son manuscritos en el idioma japonés, y expresiones matemáticas, que tuvieron que ser convertidas a lenguaje MathMl, utilizar OCR de tipo Matemático, entre otras estrategias para tal efecto, obteniéndose unos primeros resultados que muestran un primer avance en este sentido.

El gran problema de la interpretación contextual de ecuaciones matemáticas en un navegador radica en los siguientes aspectos:

- Lectura de la ecuación: Los navegadores no han sido desarrollados para trabajar con la gramática de las ecuaciones.

- interpretación del contexto: Las ecuaciones como tal carecen de significado por si solas, ellas necesitan de un texto complementario para que tengan sentido y habitualmente los navegadores no han sido diseñados para ello.
- Hacer la consulta web: los resultados de las consultas web no son las mejores, debido al alto volumen de información en internet, si la consulta no está contextualizada no se obtienen buenos resultados.

Estos aspectos poco o nada se han trabajado, pues siempre los desarrollos que se han hecho (ver tabla 1), giran alrededor del tratamiento de la ecuación como una imagen, pero no desde una perspectiva de navegación contextual Web que incluya simbología matemática, la necesidad real del usuario y el uso de algún lenguaje de consulta y elementos de inteligencia artificial, trayendo como consecuencia mejores resultados en la consulta especializada de tipo matemático. Para efectos de esta investigación se pretende desarrollar un prototipo de buscador que usa Látex como lenguaje editor de expresiones matemáticas y se observa que la calidad de la consulta se refina sustancialmente.

En consecuencia, con los motores de búsqueda la mayoría de las expresiones o son totalmente excluidas de la consulta o

solamente una fracción de estos términos matemáticos son indexados y recuperados, basta con hacer una prueba comparativa de los distintos buscadores para verificarlo.

Lo anterior al desarrollo de este trabajo de investigación para optar por el título de Magister en Ingeniería de Sistemas y Computación.

UNIVERSIDAD SIMON BOLIVAR
BIBLIOTECA
INSTITUTO DE POSTGRADO

CAPÍTULO 2.

2.1. Metodología

Para desarrollar la propuesta de este trabajo se hace necesario describir la Metodología de Investigación que se utilizará, con el objetivo de lograr resultados confiables que permitan facilitar su análisis. Por tal razón se explicará el tipo de diseño de investigación, la población, la muestra, las técnicas e instrumentos seleccionados para la recolección y análisis de los datos obtenidos.

2.1.1. Tipo de Investigación.

De acuerdo a lo mencionado en el marco teórico, es poco lo que se ha trabajado la parte de navegación contextual web de expresiones matemáticas teniendo en cuenta las preferencias del usuario, por ello, se plantea un primer prototipo de Navegador Web Contextual de este tipo, para analizar el comportamiento en la red, identificar el tipo de resultados que arroja, la efectividad de los algoritmos empleados y por ende, se abordan fuentes tanto primarias como secundarias.

2.1.2. Diseño de Investigación.

El tipo de investigación establecida es diseño Experimental - de tipo documental, esto obedece a que la investigación se sustenta en la revisión de documentos que inicialmente nos ubica en torno a las investigaciones en el campo, de tal manera que a partir de las "fisuras" encontradas dan origen a esta investigación.

Debido a las características del objeto de investigación, se trabajará con la investigación experimental - inductiva controlada, empleando un análisis dinámico debido a que se desea observar el comportamiento de un prototipo de software al realizar consultas usando expresiones de cálculo diferencial con elementos contextuales para medir la efectividad de la consulta en internet y contrastarlo con 9 de los buscadores más populares de internet.

Lo que se pretende es observar los resultados de consulta web con el navegador ontológico y compararlo con uno tradicional, comprobar la efectividad de la consulta y verificar como a partir de LATEX como herramienta de traducción de expresiones matemáticas se comporta tanto en NOMAT como en los otros navegadores.

2.1.3. Técnicas para el procesamiento de Datos.

Se diseñaron consultas de expresiones matemáticas de distinta manera y de diferente grado de complejidad, inicialmente para ver el comportamiento de cada uno de los buscadores y se comparan los resultados obtenidos en cada buscador. A continuación se observa la tabla de contrastación del comportamiento de los distintos buscadores de internet comparado con NOMAT. (Ver gráfico en la siguiente página).

Como puede observarse, los resultados que se obtienen en cada uno de los buscadores en la medida que la expresión matemática va complejizándose en la escritura los resultados comienzan a variar en cada uno de los utilizados. Al llegar a una expresión compleja tal como: $\int e^{x^2} dx$ ninguno de los buscadores comerciales, ofrece una respuesta, porque no hay forma de escribirla a menos que el usuario que haga la consulta maneje LATEX de manera óptima, debido a que todos los buscadores de internet permiten hacer la consulta directamente con ese editor de ecuaciones. Por ende NOMAT, comienza a acercarse a la condición ideal de la consulta, debido a que los resultados obtenidos son exclusivamente matemáticos y no se alejan tanto del objetivo por el cual fue empleada la herramienta de búsqueda de información.

CONTRASTACIÓN DE COMPORTAMIENTOS DE CONSULTAS DE EXPRESIONES MATEMÁTICAS EN LOS BUSCADORES MAS POPULARES DE INTERNET A PARTIR DE EXPRESIONES MATEMÁTICAS					
BUSCADORES	2X+3Y = 1	Gráfico de 2x+3y = 1	ax^2+bx+c=0	$\frac{d^2y}{dx^2}+5\frac{dy}{dx}+2y=e^{3t}$	$\int e^{x^2}dx$
Google					
YAHOO!					
AOL					
bing					
excite					

Tabla 7. Propuesta de Contrastación del comportamiento de los Buscadores comerciales mas Importantes en Internet a partir de Expresiones Matemáticas. Propuesta por el investigador.

CAPÍTULO 3.

3.1. Descripción Técnica de NOMAT

Para realizar el análisis del aplicativo NOMAT, se decidió seguir las pautas definidas mediante el lenguaje unificado de modelado UML (por sus siglas en inglés), este lenguaje nos ofrece una variedad de diagramas de modelamiento que nos ayudan a comprender mejor la complejidad del sistema debido a que permite plasmar en un lenguaje estandarizado las funcionalidades y requerimientos para lograrlas llevar a cabo en el sistema a desarrollar.

Aunque UML define diferentes tipos de diagramas para representar los aspectos del desarrollo de la aplicación, en el aplicativo NOMAT se centró en un diagrama en específico denominado diagrama de clases.

3.2. Diagrama de Clases

Un diagrama de clases es un tipo de diagrama estático que describe la estructura de un sistema mostrando sus clases, atributos y las relaciones entre ellos. Estos diagramas de clases son utilizados durante el proceso de análisis y diseño de los sistemas, donde se crea el diseño conceptual de la

información que se manejará en el sistema, y los componentes que se encargarán del funcionamiento y la relación entre uno y otro.

Para la elaboración del diagrama de clases se ha utilizado una herramienta llamada ArgoUML, permitiendo crear el diagrama del aplicativo de una forma sencilla y óptima. (Ver Gráfico)

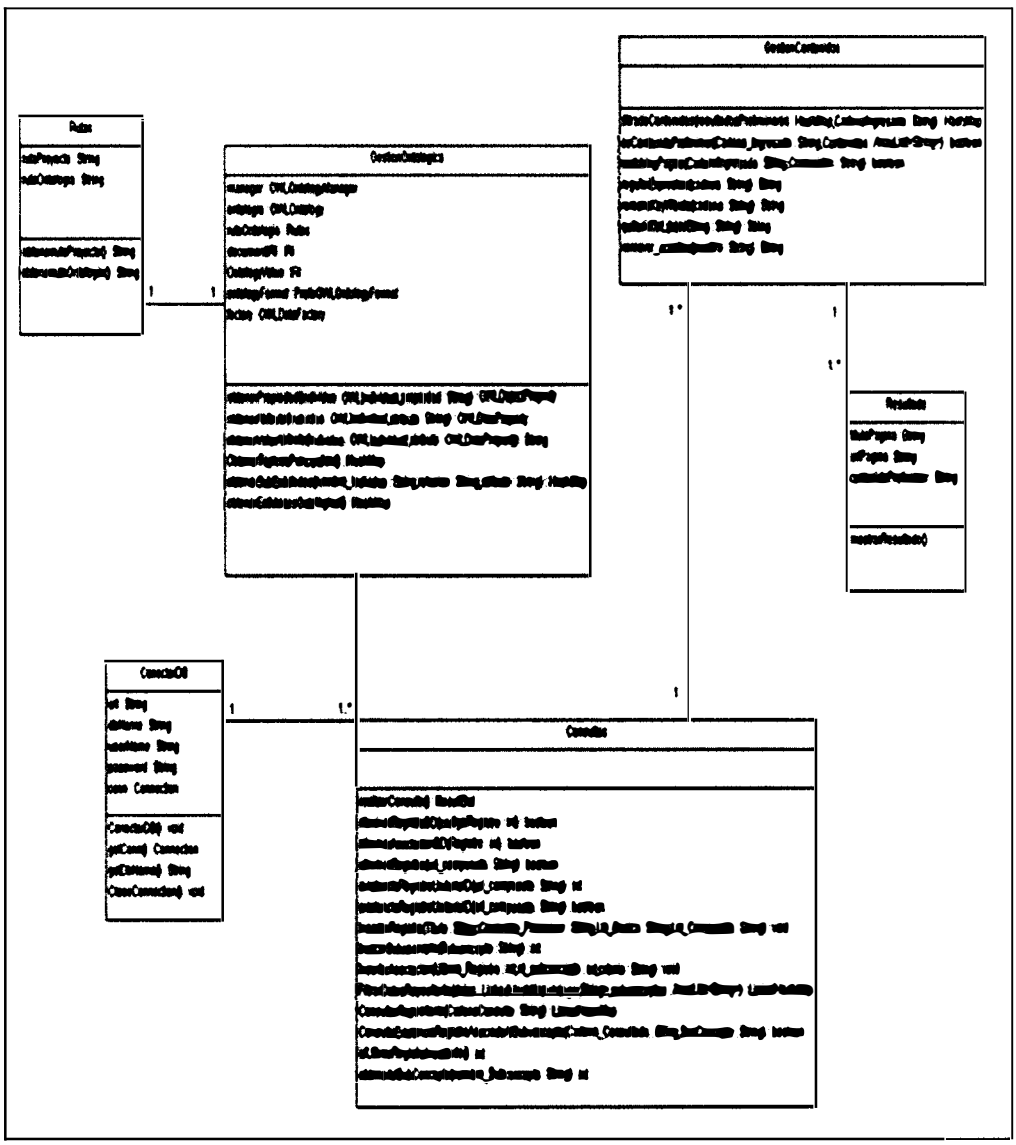


Gráfico 5. Diagrama de Clases de NOMAT

3.3. Diccionario de Clases

Un diccionario de clases es un catálogo, un depósito, de los elementos en un sistema. Como su nombre lo sugiere, estos elementos se centran alrededor de los datos y la forma en que están estructurados para satisfacer los requerimientos de los usuarios y las necesidades de la organización. Los analistas utilizan estos diccionarios por cinco razones importantes:

1. Para manejar los detalles en sistemas grandes.

2. Para comunicar un significado común para todos los elementos del sistema.

3. Para documentar las características del sistema.

4. Para facilitar el análisis de los detalles con la finalidad de evaluar las características y determinar dónde efectuar cambios en el sistema.

5. Localizar errores y omisiones en el sistema.

A continuación se describirá el diccionario de clases del aplicativo NOMAT con las clases y funciones más relevantes (Ver Tablas)arrojadas por el sistema.

.

Clase GestionOntologia	
Atributos:	**Tipo:**
Manager	OWLOntologyManager
rutaOntologia	Rutas
Ontología	OWLOntology
documentIRI	IRI
OntologyValue	IRI
ontologyFormat	PrefixOWLOntologyFormat
Factory	OWLDataFactory
Descripción de la clase	
Esta clase realiza el proceso de recorrer la ontología de NOMAT con el fin de extraer que tópicos y subtópicos del área de las matemáticas tendrá en cuenta el software para buscar , encontrar y almacenar contenidos de páginas web relacionados a dichos conceptos	

Tabla 8. Clase Gestionontologia

Método # 1 GestionOntologia	
Nombre del método:	
obtenerTopicosPrincipales	
Parámetros de Entrada:	**ninguno**
Tipo de Método:	
función	
Valor de Retorno:	
hashmap	
Descripción del Método	
Retorna todos los tópicos contenidos en la ontología de NOMAT	

Tabla9. Método # 1 Gestionontología

Método # 2 GestionOntologia	
Nombre del método:	
obtenerSubEntidades	
Parámetros de Entrada:	**Tipo:**
nombre_individuo	String
Relación	String
atributo	String
Tipo de Método:	
Función	
Valor de Retorno:	
Hashmap	
Descripción del Método	
Retorna todos los subtópicos de cada tópico contenido en la ontología de NOMAT	

Tabla 10. Método # 2 Gestionontología

Clase GestionContenidos	
Atributos:	**ninguno**
Descripción de la clase	
Esta clase realiza el proceso de filtrar los contenidos de las paginas encontradas del software NOMAT, en esta clase se realizan las comparaciones y e filtrado los contenidos de cada una de las paginas encontradas sobre la cadena de búsqueda suministrada en lenguaje látex por el usuario en el menú de búsqueda.	

Tabla 11. Clase GestionContenidos

Método # 1 GestionContenidos	
Nombre del método: quitarHTML	
Parámetros de Entrada:	**Tipo:**
cadena	String
Tipo de Método:	
función	
Valor de Retorno:	
hashmap	
Descripción del Método	
Esta función permite remover palabras de texto HTML dentro de la información capturada que son irrelevantes para la comparación de contenidos. Aquí se filtra las etiquetas HTML que tiene toda página web en los contenidos encontrados antes de buscar similitudes para nuestra cadena de texto en látex suministrada	

Tabla 12. Método # 1 GestionContenidos

Método # 2 GestionContenidos	
Nombre del método: removeKeyWords	
Parámetros de Entrada:	**Tipo:**
cadena	String
Tipo de Método:	
función	
Valor de Retorno:	
hashmap	
Descripción del Método	
Esta función permite remover cadenas de texto dentro de la información capturada que son irrelevantes para la comparación de contenidos. Aquí se filtra los comandos de colores, estilos de letra y aspecto visual de los contenidos encontrados antes de buscar similitudes para nuestra cadena de texto en látex suministrada	

Tabla 13. Método # 2 GestionContenidos

UNIVERSIDAD SIMON BOLIVAR
BIBLIOTECA
INSTITUTO DE POSTGRADO

Método # 3 GestionContenidos	
Nombre del método:	
filtradoContenidos	
Parámetros de Entrada:	**Tipo:**
resultadosPreliminares	HashMap
CadenaIngresada	String
Tipo de Método:	
función	
Valor de Retorno:	
hashmap	
Descripción del Método	
Esta función permite realizar la lógica comparativa para la determinación de coincidencias de cadenas con un alto grado de similitud para la obtención de la página web donde se encuentra el contenido de la búsqueda realizada. Cabe destacar que esta función se apoya de otras para hacer el filtrado de contenidos	

Tabla 14. Método # 3 GestionContenidos

Método # 4 GestionContenidos	
Nombre del método: esContenidoPreliminar	
Parámetros de Entrada:	**Tipo:**
Cadena Ingresada	String
Contenidos	ArrayList<String>
Tipo de Método:	
función	
Valor de Retorno:	
boolean	
Descripción del Método	
Esta función permite recorrer los contenidos encontrados de una determinada página con el fin de encontrar la cadena en látex suministrada por el usuario y determinar si la página contiene dicha cadena con el fin de capturar ese resultado	

Tabla 15. Método # 4 GestionContenidos

Clase Consultas	
Atributos:	**ninguno**
Descripción de la clase	
Esta clase realiza acciones de consultar los tópicos y subtopicos en el repositorio de información, realiza consultas a la base de datos con el fin de revisar contenidos de páginas ya consultados , como también realiza las inserciones y asociaciones en la bases de datos de NOMAT	

Tabla 16. Clase Consultas

Método # 1 Consultas	
Nombre del método: **InsertarRegistro**	
Parámetros de Entrada:	**Tipo:**
Titulo	String
Contenido_Preliminar	String
Url_Basica	String
Url_Compuesta	String
Tipo de Método:	
procedimiento	
Valor de Retorno:	
ninguno	
Descripción del Método	
Esta procedimiento permite almacenar un resultado obtenido de una consulta realiza por medio de la cadena de texto en látex suministrada por el usuario	

Tabla 17. Método # 1 Consultas

Método # 2 Consultas	
Nombre del método: ConsultarRepositorio	
Parámetros de Entrada:	**Tipo:**
cadenaConsulta	String
Tipo de Método:	
Función	
Valor de Retorno:	
LinkedHashMap<Integer,String>	
Descripción del Método	
Esta función permite consultar directamente la base de datos que contiene páginas web relacionadas a una consulta anterior, sin la necesidad de consultar en la web nuevas páginas con la información capturada	

Tabla 18. Método #2 Consultas

3.4 Diseño del Sistema

El diseño es el proceso mediante el cual se extienden, refinan y reorganizan los aspectos detectados en el proceso de modelado conceptual para generar una especificación rigurosa del sistema de información siempre orientada a la obtención de la solución del sistema software.

Además de considerar los diagramas expuestos con anterioridad, muchos son los aspectos que fueron tenidos en cuenta a la hora de Diseñar, modelar e implementar NOMAT.

Entre ellos se encuentra un Diseño Ontológico, Diseño de base de datos e interfaz gráfica requerida en el aplicativo de la presente investigación.

3.4.1. Diseño de la Ontología

La ontología aplicada para gestionar el conocimiento del sistema se conforma básicamente en un dominio de conocimiento matemático, la cual consta de una jerarquía distribuida de la siguiente manera (ver gráfico):

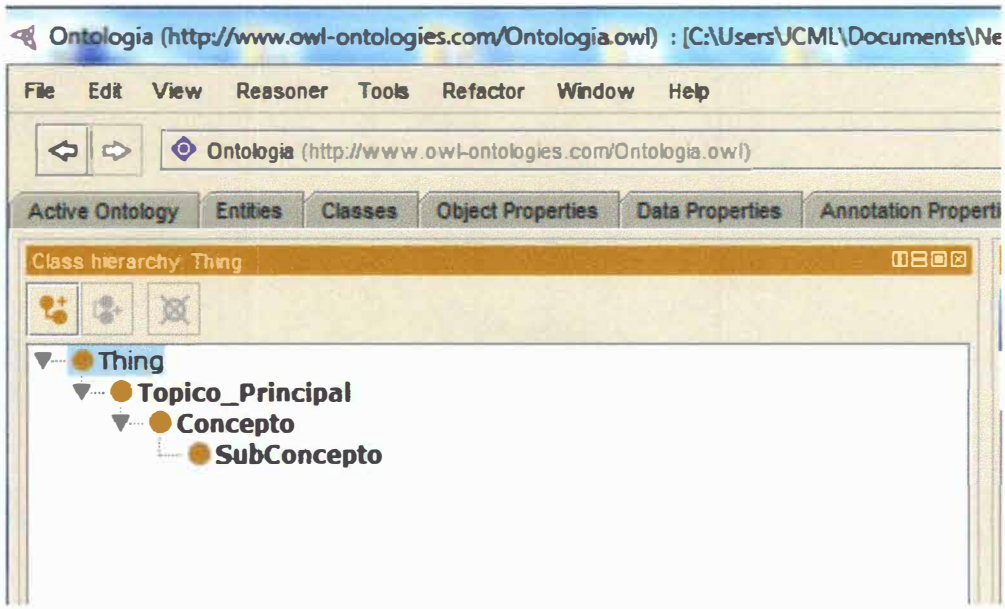


Gráfico 6. Arbol Jerárquico de la Ontología Empleada en NOMAT. Tomado del pantallazo de la herramienta Protegé

Existe una clase principal llamada **tópico_principal**, compuesto de 3 tópicos matemáticos llamados (cálculo diferencial, cálculo integral y cálculo vectorial), esta clase se subdivide con otra denominada **Concepto** los cuales contienen 3 individuos que pertenecen a un denominado tópico, y estos conceptos a la vez se subdividen en una clase llamada **Subconcepto** la cual contiene 24 individuos que pertenecen a su vez a un concepto asociado.

La siguiente ilustración muestra las clases y cada individuo que conforma el diseño ontológico de la aplicación NOMAT. Así mismo se puede observar qué relaciones tiene cada individuo de una clase con respecto a otro individuo perteneciente a otra clase.

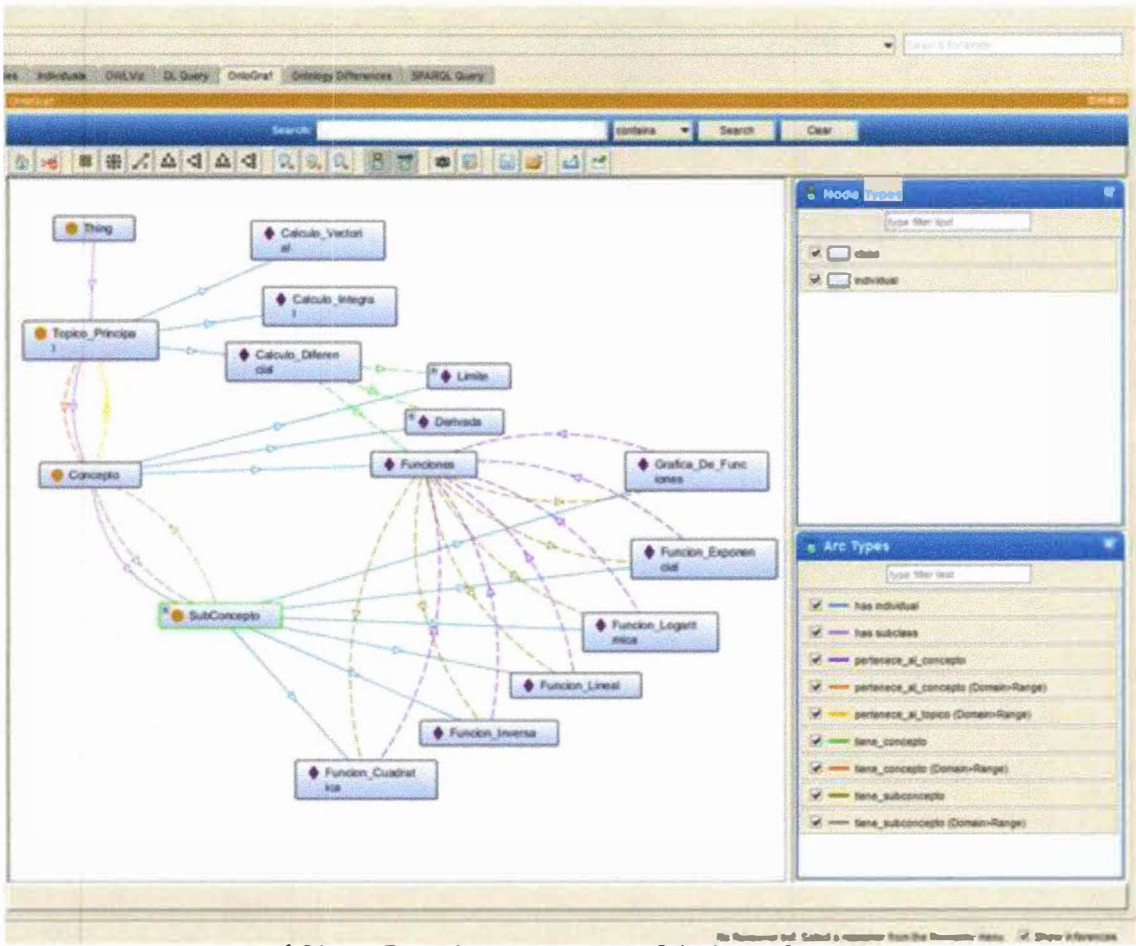


Gráfico 7. Diagrama Ontológico de NOMAT

Imagen capturada del Protegé

3.4.2. Diseño de la Base de Datos

Una base de datos correctamente diseñada permite obtener acceso a información exacta y actualizada. La base de datos del aplicativo NOMAT permite almacenar los resultados de búsqueda exitosos generados por las acciones de consulta, con

el fin de recuperarlos de manera ágil cuando se necesite volverlos a consultar.

Toda base de datos posee una estructura de organización mediante tablas conformadas por registros y campos, para la implementación y recuperación de los resultados de NOMAT fueron conformadas 3 tablas base llamadas **Subconceptos**, **asociacion_registros** y **registros**.

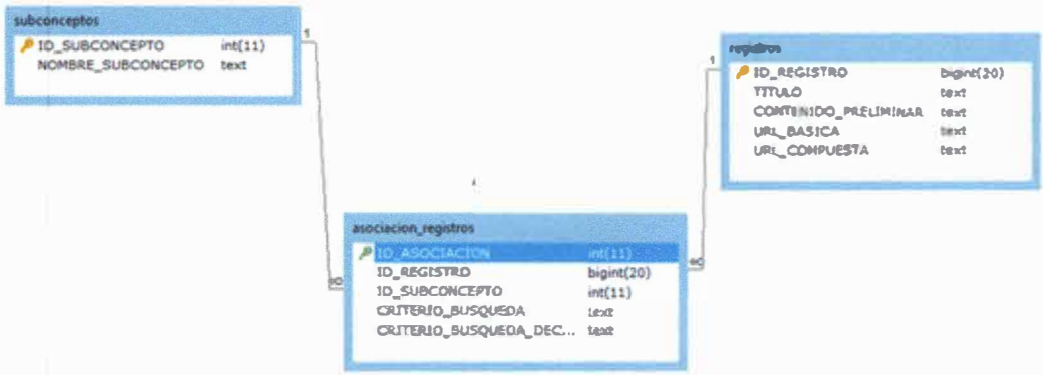


Gráfico 8. Diagrama de la Base de Datos de NOMAT

La tabla **subconcepto** almacena el nombre y el código de cada uno de los individuos de la clase subconceptos de la ontología.

La tabla **asociacion_registros** almacena la cadena suministrada por el usuario de un resultado exitoso guardado por NOMAT.

La tabla **registros** almacena, el título de la página, el contenido preliminar, la url raíz de la página y la url compuesta de la misma de resultados exitosos encontrados por NOMAT.

3.4.3. Diseño de la Interfaz Gráfica

- **Interfaz Gráfica Principal**

El aplicativo NOMAT posee una sola ventana para la interacción donde está conformada por los siguientes elementos (ver gráfico):

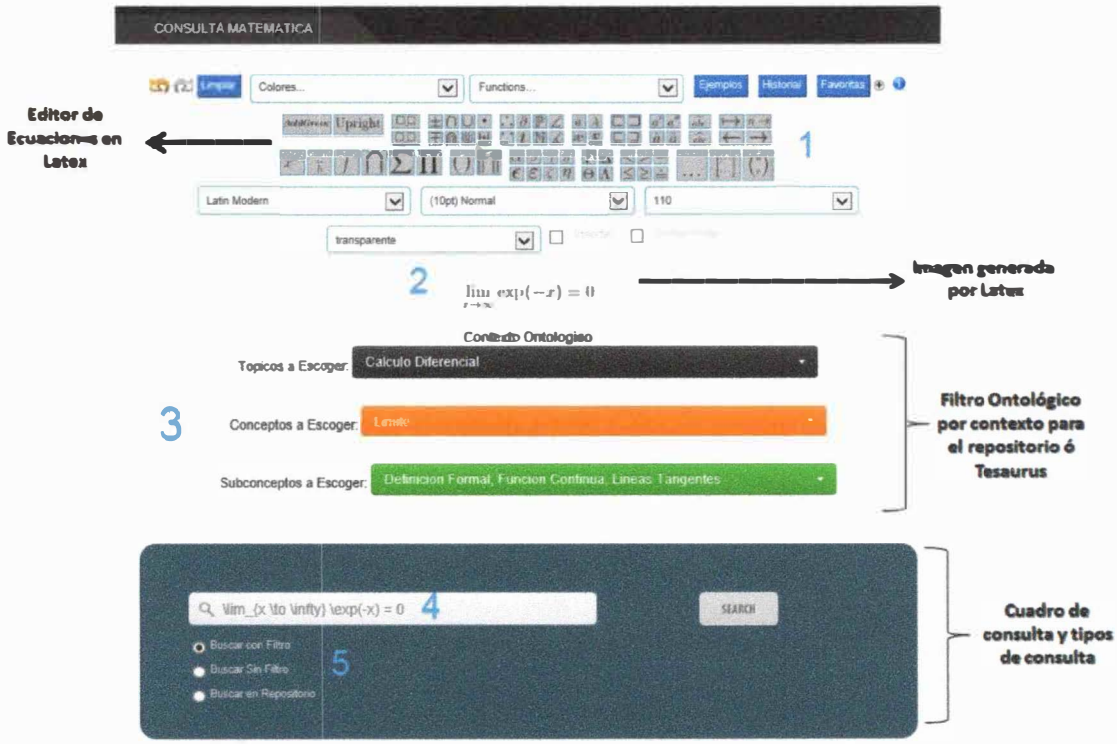


Gráfico 9. Interfaz gráfica del Navegador NOMAT

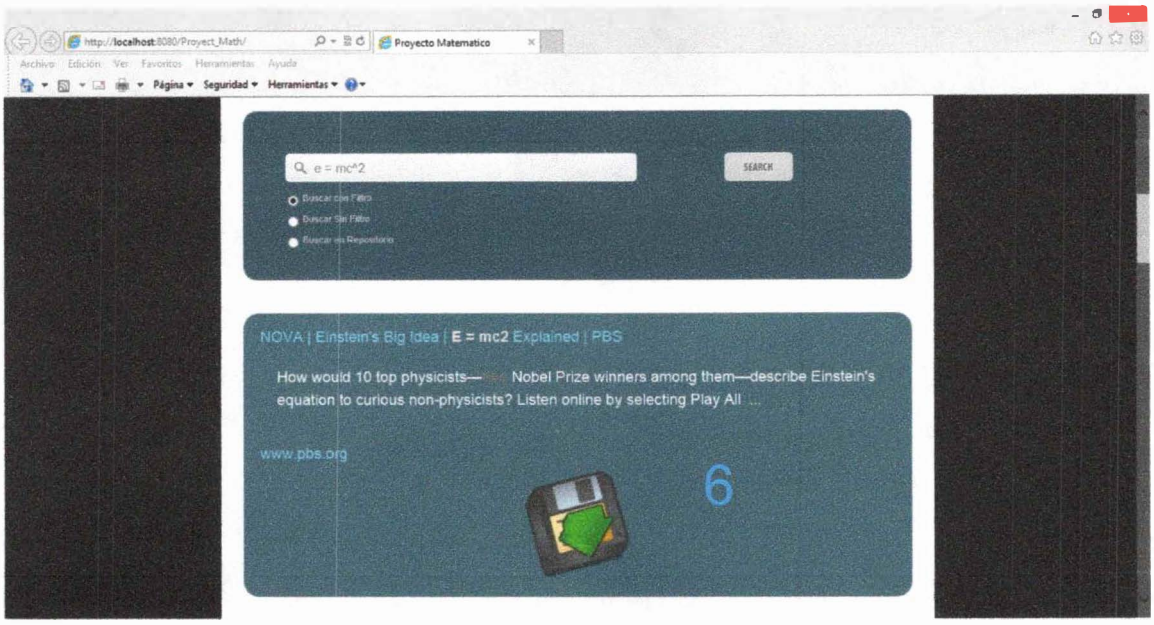


Gráfico 10. Panel de Resultados de Consulta

1. Un editor de ecuaciones para formato látex denominado Editor de ecuaciones CodeCogs. Este editor nos permite crear ecuaciones matemáticas en cualquier sitio web para ser consultada para la búsqueda de páginas que contenga dicha ecuación.
2. Un visor de la cadena que se desea buscar. Este visor traduce el código látex para mostrarnos la cadena que se intenta buscar.
3. Un Contexto Ontológico que permite ir clasificando las páginas web relacionadas con la búsqueda a efectuar.
4. Un campo de texto para la búsqueda del texto suministrado por el editor matemático.

5. Un conjunto de opciones de búsqueda que le permiten al usuario seleccionar el repositorio que se desea buscar o poder refinar la búsqueda.
6. Un panel de resultados que describe todas las páginas que contienen la cadena suministrada, destacando la dirección completa, la dirección raíz y un contenido preliminar de la página.

3.5. Implementación y Desarrollo.

Para el Desarrollo e Implementación se tuvo en cuenta el manejo de las siguientes tecnologías:

3.5.1. Interfaz De Usuario

- **XML Y HTML**

XML es un sistema estándar de codificación de información. Los programas que utilizan el formato XML pueden intercambiar fácilmente sus datos, ya que responden a una misma lógica interna.

Los documentos XML son ficheros de texto que contienen la información organizada en forma de árbol: cada rama puede tener unos atributos propios y servir de base para otras ramas. Además, los documentos XML se pueden transformar (por ejemplo,

a formato HTML, para mostrar la información en una página web), o combinar: un tronco con todas sus ramas puede pasar a ser una rama de otro árbol mayor.

Varias son las características que ofrece XML: Aunque hoy día XML aún no está tan extendido como HTML, su uso futuro en la Web mejorará la eficiencia de las búsquedas, al proporcionar cada documento XML metadatos sobre sí mismo.

Permite proporcionar diferentes vistas sobre los datos (HTML, PDF, voz, etc.), dependiendo de quién sea el cliente.

Facilita la integración desde fuentes de datos heterogéneas, por ejemplo, páginas Web, distintas bases de datos,...

Los documentos tienen una estructura que los hace legibles e inteligibles no sólo para los ordenadores, sino también para los humanos.

Las aplicaciones de XML son fácilmente extensibles mediante definiciones de nuevos tipos de documento

HTML son las siglas de HyperText Markup Language (Lenguaje de Marcado de Hipertexto), es el lenguaje de marcado predominante para la elaboración de páginas web. Es usado para describir la estructura y el contenido en forma de texto, así como para complementar el texto con objetos tales como

UNIVERSIDAD SIMON BOLIVAR
BIBLIOTECA
INSTITUTO DE POSTGRADO

imágenes. HTML se escribe en forma de "etiquetas", rodeadas por corchetes angulares (<,>).

El lenguaje HTML es un estándar reconocido en todo el mundo y cuyas normas define un organismo sin ánimo de lucro llamado World Wide Web Consortium, más conocido como W3C.

Como se trata de un estándar reconocido por todas las empresas relacionadas con el mundo de Internet, una misma página HTML se visualiza de forma muy similar en cualquier navegador de cualquier sistema operativo. El propio W3C define el lenguaje HTML como "un lenguaje reconocido universalmente y que permite publicar información de forma global". Por convención, los archivos de formato HTML usan la extensión .htm o .HTML.

- **CSS**

Las hojas de estilo en cascada (Cascading Style Sheets, CSS) son un lenguaje formal usado para definir la presentación de un documento estructurado escrito en HTML o XML (y por extensión en XHTML). El W3C es el encargado de formular la especificación de las hojas de estilo que servirá de estándar para los agentes de usuario o navegadores.

La idea que se encuentra detrás del desarrollo de CSS es separar la estructura de un documento de su presentación. La

información de estilo puede ser adjuntada tanto como un documento separado o en el mismo documento HTML.

En este último podrían definirse estilos generales en la cabecera del documento o en cada etiqueta particular mediante el atributo "style".

Las ventajas de utilizar CSS (u otro lenguaje de estilo) son:

- Control centralizado de la presentación de un sitio web completo, con lo que se agiliza de forma considerable la actualización del mismo.

- Los navegadores permiten a los usuarios especificar su propia hoja de estilo local que será aplicada a un sitio web remoto, con lo que aumenta considerablemente la accesibilidad.

- Una página puede disponer de diferentes hojas de estilo según el dispositivo que la muestre.

- El documento HTML en sí mismo es más claro de entender y se consigue reducir considerablemente su tamaño.

3.5.2. Controles

- **JavaScript**

JavaScript es un lenguaje interpretado utilizado principalmente en páginas web, con una sintaxis semejante a la del lenguaje Java. Sin embargo, al contrario que Java, JavaScript no es un lenguaje orientado a objetos propiamente dicho, ya que no dispone de herencia. Es más bien un lenguaje basado en prototipos, ya que las nuevas clases se generan clonando las clases base (prototipos) y extendiendo su funcionalidad.

Todos los navegadores interpretan el código JavaScript integrado dentro de las páginas web. Para interactuar con una página web se provee al lenguaje JavaScript de una implementación del DOM (Modelo de Objetos del Documento). JavaScript se ejecuta en el agente de usuario al mismo tiempo que las sentencias van descargándose junto con el código HTML.

- **JQuery**

jQuery es una biblioteca o framework de JavaScript que permite simplificar la manera de interactuar con los documentos HTML, manipular el árbol DOM, manejar eventos, desarrollar

animaciones y agregar interacción con la tecnología AJAX a páginas web.

jQuery, al igual que otras bibliotecas, ofrece una serie de funcionalidades basadas en JavaScript que de otra manera requerirían de mucho más código, es decir, con las funciones propias de esta biblioteca se logran grandes resultados en menos tiempo y espacio.

- **Ajax**

AJAX, acrónimo de Asynchronous JavaScript And XML (JavaScript asíncrono y XML), es una técnica de desarrollo web para crear aplicaciones interactivas o RIA (Rich Internet Applications). Estas aplicaciones se ejecutan en el cliente, es decir, en el navegador de los usuarios mientras se mantiene la comunicación asíncrona con el servidor en segundo plano.

De esta forma es posible realizar cambios sobre las páginas sin necesidad de recargarlas, lo que significa aumentar la interactividad, velocidad y usabilidad en las aplicaciones.

- **Servlet**

Entre las muchas particularidades que distinguen a la Web de los restantes medios de comunicación, está la capacidad de interacción. Interacción que inicialmente era bastante reducida y se conseguía a través de código HTML y algunos elementos embebidos de JavaScript. Cuando se mencionaba el nombre de Java en este entorno, algunos lo asociaban a JavaScript, y otros, un poco más conocedores, lo asociaban con pequeños programas que se ejecutaban en el cliente y se llamaban Applets. Esto realmente fue el inicio de Java en la Web y surgió para darle más vida y dinamismo a la misma.

Los Servlet son la respuesta de la tecnología en Java a la programación de la Interfaz de Compuerta Común (CGI). Son programas que se ejecutan en el servidor, realizando la función de una capa intermedia entre una petición proveniente de un navegador Web u otro cliente HTTP, y las aplicaciones del servidor, pudiendo utilizar toda la paquetería y potencialidades del lenguaje. Su función principal es proveer páginas web dinámicas y personalizadas, utilizando para este objetivo los accesos a bases de datos, flujos de trabajo y otros recursos.

- **JSP**

JSP es un acrónimo de Java Server Pages, que en castellano vendría a decir algo como Páginas de Servidor Java. Es, pues, una tecnología orientada a crear páginas web con programación en Java.

Con JSP podemos crear aplicaciones web que se ejecuten en variados servidores web, de múltiples plataformas, ya que Java es en esencia un lenguaje multiplataforma.

Las páginas JSP están compuestas de código HTML/XML mezclado con etiquetas especiales para programar scripts de servidor en sintaxis Java. Por tanto, las JSP podremos escribirlas con nuestro editor HTML/XML habitual.

El motor de las páginas JSP está basado en los servlets de Java -programas en Java destinados a ejecutarse en el servidor, aunque el número de desarrolladores que pueden afrontar la programación de JSP es mucho mayor, dado que resulta mucho más sencillo aprender que los servlets.

En JSP creamos páginas de manera parecida a como se crean en ASP o PHP -otras dos tecnologías de servidor-. Generamos archivos con extensión .jsp que incluyen, dentro de la estructura de etiquetas HTML, las sentencias Java a ejecutar

en el servidor. Antes de que sean funcionales los archivos, el motor JSP lleva a cabo una fase de traducción de esa página en un Servet, implementado en un archivo class (Byte codes de Java). Esta fase de traducción se lleva a cabo habitualmente cuando se recibe la primera solicitud de la página .jsp, aunque existe la opción de precompilar en código para evitar ese tiempo de espera la primera vez que un cliente solicita la página.

3.5.3. Modelo

- **Java**

Java es un lenguaje de programación con el que podemos realizar cualquier tipo de programa. En la actualidad está muy extendido y cada vez cobra más importancia tanto en el ámbito de Internet como en la informática en general. Fué desarrollado por la compañía Sun Microsystems con gran dedicación y siempre enfocado a cubrir las necesidades tecnológicas más punteras.

Una de las principales características por las que Java se ha hecho muy famoso, es que es un lenguaje independiente de la plataforma. Eso quiere decir que si hacemos un programa en Java podrá funcionar en cualquier ordenador del mercado. Es una ventaja significativa para los desarrolladores de software,

pues antes tenían que hacer un programa para cada sistema operativo, por ejemplo Windows, Linux, Apple, etc. Esto lo consigue porque se ha creado una Máquina de Java para cada sistema que hace de puente entre el sistema operativo y el programa de Java y posibilita que este último se entienda perfectamente.

La independencia de plataforma es una de las razones por las que Java es interesante para Internet, ya que muchas personas deben tener acceso con ordenadores distintos. Pero no se queda ahí, Java está desarrollándose incluso para distintos tipos de dispositivos además del ordenador como móviles, agendas y en general para cualquier cosa que se le ocurra a la industria.

Java fue pensado originalmente para utilizarse en cualquier tipo de electrodoméstico pero la idea fracasó. Uno de los fundadores de Sun rescató la idea para utilizarla en el ámbito de Internet y convirtieron a Java en un lenguaje potente, seguro y universal gracias a que lo puede utilizar todo el mundo y es gratuito. Una de los primeros triunfos de Java fue que se integró en el navegador Netscape y permitía ejecutar programas dentro de una página web, hasta entonces impensable con el HTML.

Actualmente Java se utiliza en un amplio abanico de posibilidades; casi cualquier cosa que se puede hacer en otros lenguajes se puede hacer también en Java y muchas veces con grandes ventajas. Para lo que interesa a la presente investigación, con Java se puede programar páginas web dinámicas, con accesos a bases de datos, utilizando XML, con cualquier tipo de conexión de red entre distintos sistemas. En general, todas las aplicaciones que se desee hacer con acceso a través de una web se puede desarrollar utilizando Java.

3.5.4. Librerías

- **Google search ajax api**

Google search AJAX API es una API experimental o «beta» (versión 0.1) creada por Google que permite a los desarrolladores Web acceder a los servicios de búsqueda de Google en sus propias páginas Web.

Se trata específicamente de una biblioteca javascript que proporciona unos objetos simples que realizan búsquedas en línea sobre un número de servicios de Google (búsqueda Web, búsqueda local, vídeo y búsqueda de Blog); todo ello a través de programación javascript y XML (AJAX).

Constituye una de las tecnologías que compone los Web service. Esto significa que el usuario puede acceder a un sitio

Web que implementa esta api, realizar una consulta y obtener los resultados de la misma en el mismo sitio Web sin saber que se está empleando un servicio Google. Éste es el concepto detrás de Web service.

Se ofrece gratuitamente para que los desarrolladores puedan probarla, ya que aún no existe ninguna versión comercial. Pero para empezar a utilizarla es necesario registrarse previamente y obtener una clave que otorga un máximo de 1000 peticiones por día. Esto a fines de evitar que otro servidor Web (que implementa servicios de búsqueda de Google) funcione paralelamente a Google.

- **Editor de ecuaciones latex en línea (CodeCogs)**

Editor de ecuaciones CodeCogs puede ser puede ser utilizado en una amplia gama de circunstancias para crear ecuaciones matemáticas muy bien formateados en látex o MATHML. El servicio ha sido optimizado cuidadosamente en los últimos 6 años y es conocido por su fiabilidad. Se ha probado a fondo por millones de usuarios y genera millones de ecuaciones cada día por miles de sitios web - lo que es # 1 editor en línea del mundo.

El editor viene en dos modos básicos:

- **Modo Standalone** - Se utiliza principalmente para crear ecuaciones que luego se copian en cualquier sistema de

UNIVERSIDAD SIMON BOLIVAR
BIBLIOTECA
INSTITUTO DE POSTGRADO

terceros, incluidos otros sitios web, blogs, programas de correo electrónico y aplicaciones de escritorio. Abrir editor independiente

- **Modo integrado** - Está diseñado para integrarse físicamente dentro de una web, que permite a sus usuarios crear ecuaciones sin moverse de su sitio web. Hay dos enfoques para la integración del editor: ya sea como una ventana emergente que se pueden activar desde un botón en su sitio, o como un objeto incrustado directamente a una página. Una gama de plugins para editores populares están también disponibles, por ejemplo, CK Editor, Tiny MCE.

Para ambas variantes del editor, el diseño se puede personalizar para ofrecer sólo los paneles desplegables que necesita y en el orden que se desee.

- **Jsoup: Java HTML Analizador**

Jsoup es una biblioteca de Java que sirve para trabajar con HTML en el mundo real. Proporciona una API muy conveniente para la extracción y manipulación de datos, utilizando lo mejor de DOM, CSS, y los métodos de jQuery similares.

Jsoup implementa la especificación WHATWG HTML5, y analiza HTML en el mismo DOM como los navegadores modernos lo hacen.

Jsoup identifica y analiza código HTML de una URL, archivo o cadena encontrar y extraer los datos, utilizando transversales DOM o selectores CSS para manipular los elementos HTML, atributos y texto contenido enviado por los usuarios

Jsoup está diseñado para hacer frente a todas las variedades de HTML que se encuentran en la naturaleza.

- **OWLAPI 3.4.4**

OWLAPI es una Java API que permite crear, manipular y serializar ontologías OWL. La última versión de la API está enfocada a OWL 3

OWLAPI es de código abierto y está disponible en régimen de LGPL o Apache Licencias

La API de OWL incluye los siguientes componentes:

- Una API para OWL 2 y una implementación de referencia en memoria eficiente
- RDF / XML parser y escritor
- OWL / XML parser y escritor
- OWL funcional Sintaxis analizador y escritor
- Turtle analizador y escritor
- KRSS parser
- OBO Flat formato de archivo parser

- Interfaces Reasoner para trabajar con razonadores como hecho , ermitaño, Pellet y Racer

La versión original de la API para OWL 1.0 fue desarrollado como parte del proyecto WonderWeb. Versión 2.0.0 del API para OWL fue desarrollado como parte de la CO-ODE proyecto y la TONOS proyecto. Versión 3.5.0 es la última versión.

- **JSON-lib**

JSON-lib es una biblioteca Java para la transformación de los fríjoles, mapas, colecciones, arrays de Java y XML a JSON y viceversa, Beans y DynaBeans.

Está basada en la obra de Douglas Crockford en http://www.json.org/java .

Las siguientes tablas repasan los tipos de la conversión entre Java y JavaScript:

JSON		Java
cadena	<=>	java.lang.String, java.lang.Character, carbón de leña
número	<=>	java.lang.Number, byte, short, int, long, float, double
true | false	<=>	java.lang.Boolean, booleano
Nulo	<=>	Nulo
función	<=>	net.sf.json.JSONFunction
colección	<=>	net.sf.json.JSONArray (Object, String, Number, Boolean, función)
objeto	<=>	net.sf.json.JSONObject

Tabla 19. Tipos de Conversión entre Java y Java Script.

Json-lib requiere (al menos) las siguientes dependencias en la ruta de clases:

- jakarta commons-lang 2.5
- jakarta Commons-BeanUtils 1.8.0
- jakarta commons-colecciones 3.2.1
- jakarta commons-logging 1.1.1
- ezmorph 1.0.6

3.6. Software de apoyo

- **Netbeans 7.4**

Netbeans es un entorno integrado de desarrollo o IDE (Integrated Development Environment), en él podemos realizar todas las tareas asociadas a la programación:

- Editar el código
- Compilarlo
- Ejecutarlo
- Depurarlo

NetBeans es un entorno de desarrollo muy completo y profesional. Contiene muchas funcionalidades, para distintos tipos de aplicaciones y para facilitar al máximo la programación, la prueba y la depuración de las aplicaciones que se desarrollan. También incorpora un editor propio.

- **Protégé 4.1**

Es un editor de código abierto usado para construir Ontologías y un marco general para representar el conocimiento. Está escrito en Java que es un lenguaje de programación orientado a objetos. Se usa para construir aplicaciones para la Web Semántica, para hacer una descripción semántica de la información. Sus archivos se hacen en el lenguaje OWL.

Protége es un editor de Ontologías cuyo sitio oficial es la Universidad de Stanford.Fúe creado por Stanford Center for BiomedicalResearch.

- **Glassfich Server 3.1.2.2**

GlassFish es un servidor de aplicaciones de software libre desarrollado por Sun Microsystems, compañía adquirida por Oracle Corporation, que implementa las tecnologías definidas en la plataforma Java EE y permite ejecutar aplicaciones que siguen esta especificación. Es gratuito, de código libre y se distribuye bajo un licenciamiento dual a través de la licencia CDDL y la GNU GPL. La versión comercial es denominada Oracle GlassFish Enterprise Server (antes Sun GlassFish Enterprise Server).

GlassFish está basado en el código fuente donado por Sun y Oracle Corporation; este último proporcionó el módulo de persistencia TopLink. GlassFish tiene como base al servidor Sun Java System Application Server de Oracle Corporation, un derivado de Apache Tomcat, y que usa un componente adicional llamado Grizzly que usa Java NIO para escalabilidad y velocidad.

CAPÍTULO 4.

4.1. Experimentación.

Para verificar el comportamiento del prototipo de buscador de expresiones matemáticas se usaron las herramientas de consulta de información más populares en Internet (Nicolás, 2014) con el objetivo de elaborar una tabla de contrastación y observar los resultados respectivos en cada uno de ellos.

A continuación se anexan algunos gráficos de distintos buscadores que muestran el respectivo comportamiento de acuerdo a las ecuaciones: (ver gráficos en la siguiente página)

Inicialmente se utilizó **Google** y se consultó por la expresión $2x + 3y = 1$ y arrojó 5.470.000 resultados 0,29 segundos. Aunque los resultados fueron en esencia de tipo matemático las páginas revisadas inicialmente asumían la expresión pero con signo negativo. En el caso de Yahoo, arrojó resultados un poco más exactos, comparados con los de Google en el sentido de la forma de la expresión consultada.

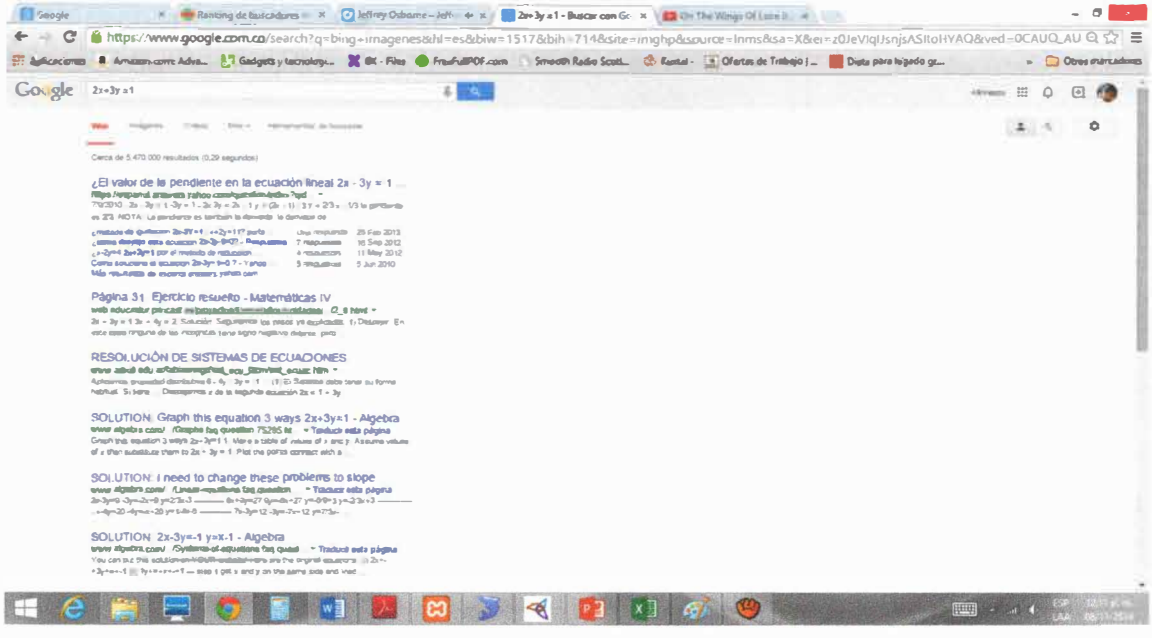


Gráfico 11. Consulta en Google de la expresión $2x+3y=1$. Tomado desde: https://www.google.com.co/?gws_rd=ssl#q=2x%2B3y%3D1

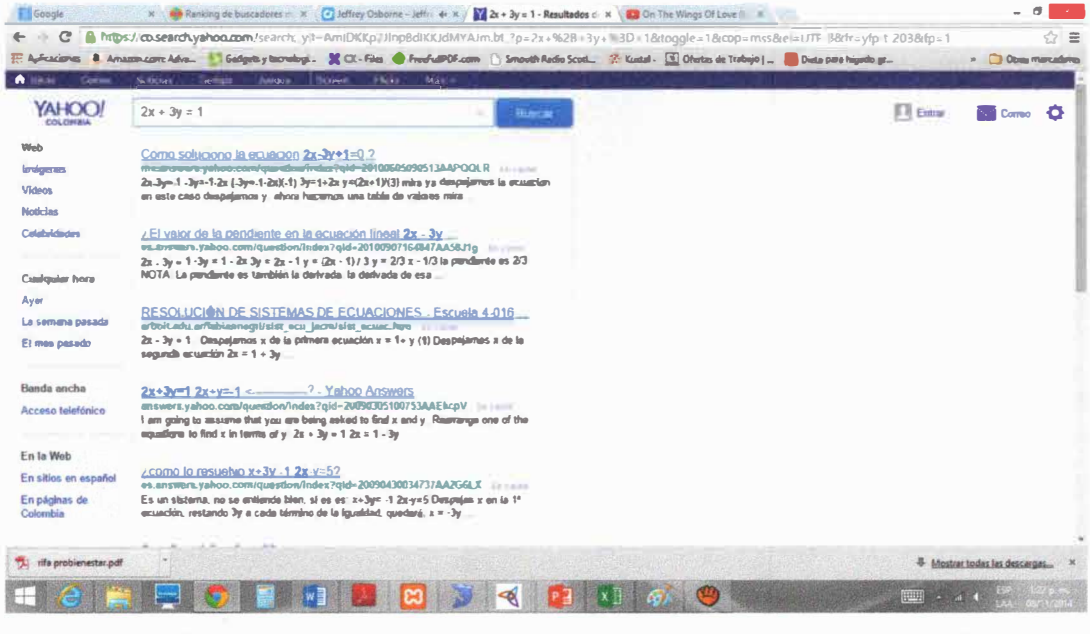


Gráfico 12. Consulta en Yahoo de la expresión $2x+3y=1$. Tomado desde: https://co.search.yahoo.com/search;_ylt=A0LEV2tvY.FU5isAIYSqcgx.;_ylc=X1MDMjExNDc0NTAwMgRfcgMyBGZyAwRncHJpZANJQ1JEemNBZ1JJaWFncmJx81JkbHBBBG5fcnNsdAMwBG5fc3VnZwMxBG9yaWdpbgNjby5zZWFyY2gueWFob28uY29tBHBvcwMwBHBxc3RyAwRwcXN0cmwDBHFzdHJsAzcEcXV1cnkDMngrM3k9MQR0X3N0bXADMTQyNDA1NzIxMA--?p=2x%2B3y%3D1&fr=sfp&fr2=sb-top-co.search&iscqry=

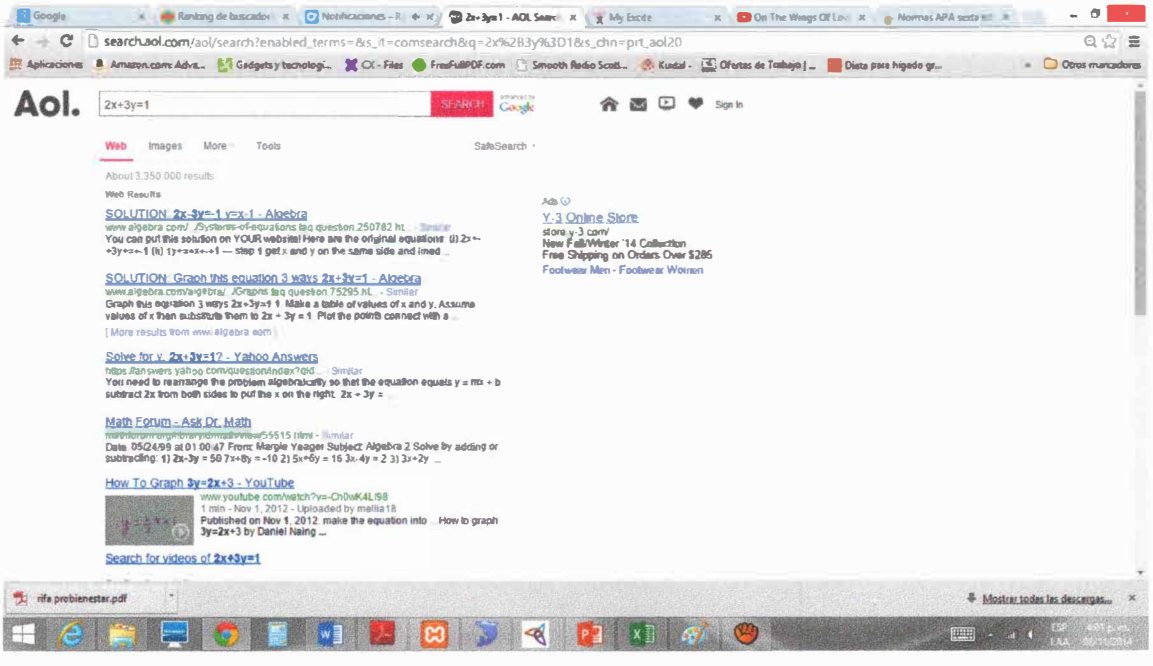


Gráfico 13. Consulta en AOL de la expresión $2x + 3y = 1$ Tomado desde: http://search.aol.com/aol/search?enabled_terms=&s_it=comsearch&q=2x%2B3y%3D1&s_chn=prt_ct11

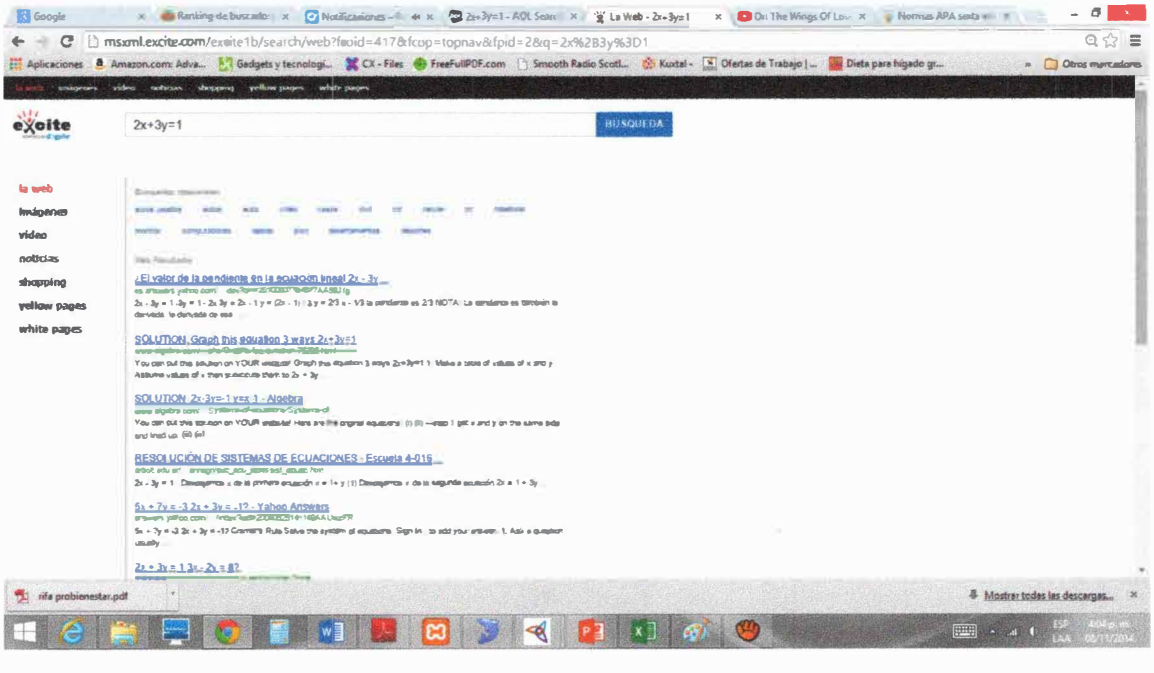


Gráfico 14. Consulta en Excite de la expresión 2x+3y=1. Tomado desde: http://msxml.excite.com/search/web?fcoid=417&fcop=topnav&fpid=2&q=2x%2B3y%3D1

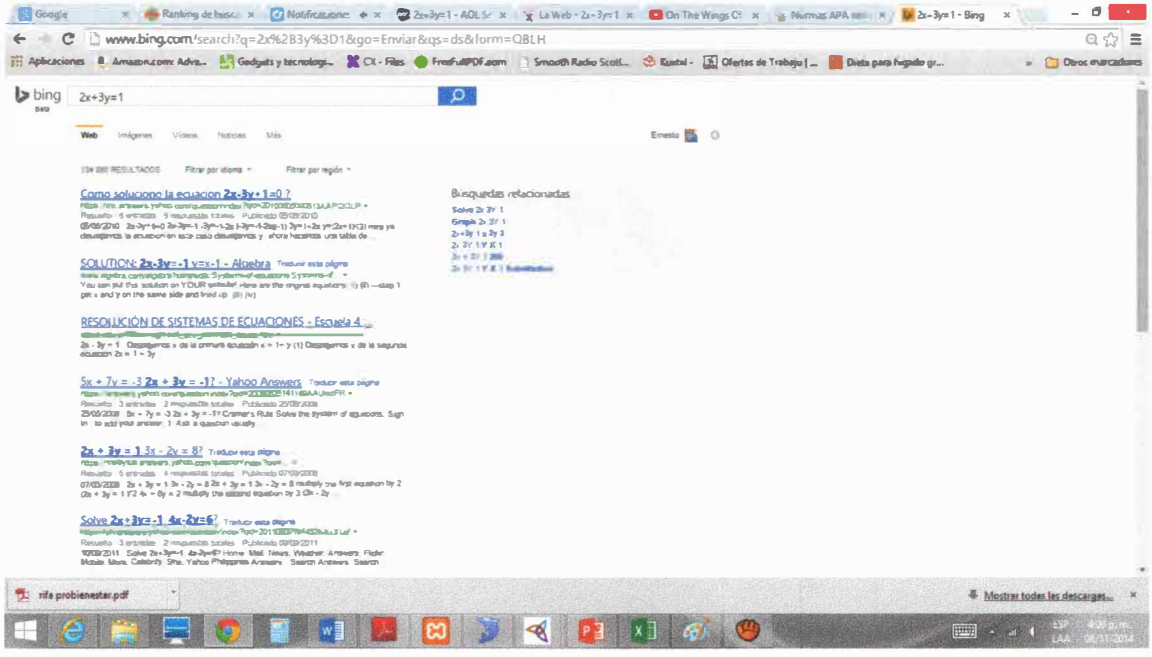


Gráfico 15. Consulta en Bing de la expresión 2x+3y=1. Tomado desde: http://www.bing.com/search?q=2x%2B3y%3D1&go=Enviar&qs=ds&form=QBLH

Esta misma consulta se hizo con las otras herramientas y los resultados fueron muy similares a los obtenidos con Google. En la medida en que las expresiones fueron subiendo el nivel de complejidad los resultados empezaron a bajar la efectividad. Por ejemplo para una ecuación diferencial por ejemplo: $\frac{d^2y}{dx^2} + 5\frac{dy}{dx} + 2y = e^{3t}$ o para una expresión con integrales tal como la que se presenta a continuación $\int e^{x^2} dx$, ninguno de los buscadores tienen la forma de ser escritas a menos que el usuario que desee hacer la consulta la escriba directamente en LATEX. A

UNIVERSIDAD SIMON BOLIVAR
BIBLIOTECA
INSTITUTO DE POSTGRADO

continuación se mostrará una tabla que contrasta los resultados obtenidos en cada buscador empleado, sin tener en cuenta el tiempo, que para efectos de esta investigación no es relevante:

CONTRASTACIÓN DE LOS COMPORTAMIENTOS DE CONSULTAS A PARTIR DE EXPRESIONES MATEMÁTICAS EN LOS BUSCADORES DE INTERNET MAS POPULARES					
Buscadores	2x+3y=1	Gráfico de 2x+3y=1	ax^2+bx+c=0	$\frac{d^2y}{dx^2} - 5\frac{dy}{dx} + 2y = e^x$	$\int e^{x^2} dx$
Google	Arroja resultados relacionados con expresiones matemáticas, incluye texto, imágenes, video, etc, sin embargo no interpreta la expresión con el mismo signo	En este caso la consulta tiene un texto complementario contextualizandola, obteniendose mejores resultados	Arroja menor número de páginas que los otros buscadores, aunque todas las mostradas son de tipo matemático	No hay manera de escribir esta ecuación, solo es posibl onsultar a partir del texto complementario	No hay manera de escribir esta ecuación, solo es posibl onsultar a partir del texto complementario
YAHOO!	Arroja resultados relacionados con expresiones matemáticas, incluye texto, imágenes, video, etc, pero yahoo encuentra páginas con mayor precisión en la forma de la ecuación.	El comportamiento es muy similar al de google	Resultados muy similares a los de google, pero con mayor número de páginas	No hay manera de escribir esta ecuación, solo es posibl onsultar a partir del texto complementario	No hay manera de escribir esta ecuación, solo es posibl onsultar a partir del texto complementario
AOL	El comportamiento es muy similar al de google	El comportamiento es muy similar al de google	Resultados muy similares a los de google, pero con mayor número de páginas	No hay manera de escribir esta ecuación, solo es posibl onsultar a partir del texto complementario	No hay manera de escribir esta ecuación, solo es posibl onsultar a partir del texto complementario
bing	El comportamiento es muy similar al de google	El comportamiento es muy similar al de google	Resultados muy similares a los de google, pero con mayor número de páginas	No hay manera de escribir esta ecuación, solo es posibl onsultar a partir del texto complementario	No hay manera de escribir esta ecuación, solo es posibl onsultar a partir del texto complementario
excite	El comportamiento es muy similar al de google	El comportamiento es muy similar al de google	Resultados muy similares a los de google, pero con mayor número de páginas	No hay manera de escribir esta ecuación, solo es posibl onsultar a partir del texto complementario	No hay manera de escribir esta ecuación, solo es posibl onsultar a partir del texto complementario

Tabla 20. Contrastación de comportamientos de consultas de expresiones matemáticas en los buscadores de internet

En lo que respecta a NOMAT, la consulta a continuación se hará una descripción técnica detallada del modelo de consulta en esta herramienta.

4.2. Modelo de Búsqueda de NOMAT

En esta sección se describe más detalladamente cada uno de los distintos componentes que conforman el aplicativo NOMAT.

A continuación se expone un esquema arquitectónico para la consulta de páginas relacionadas mediante cadenas de búsqueda enriquecida por texto látex, el cual engloba a cada componente, sus interacciones y resultados dentro de sus procesos:

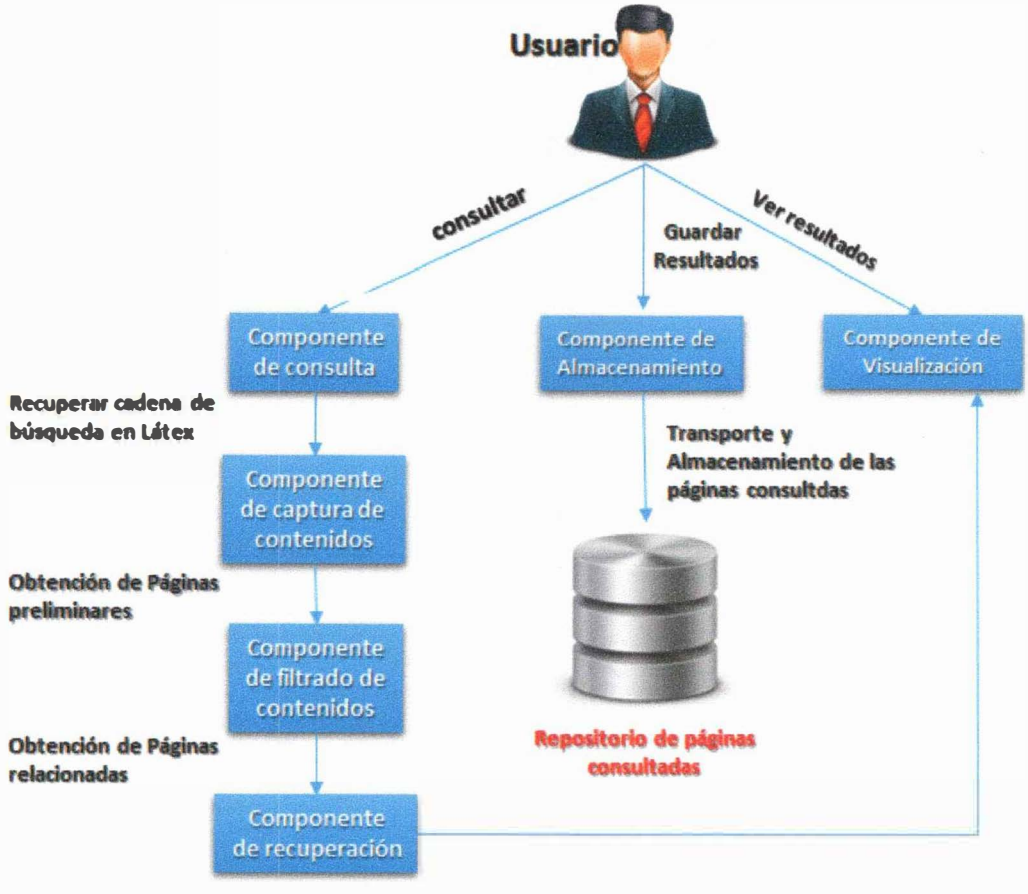


Gráfico 16. Arquitectura de NOMAT.

Fuente: gráfico creado por el investigador

4.2.1. Componente de Consulta

El componente de consulta es el responsable de proporcionarle al usuario una interfaz gráfica principal, este consta del editor matemático CodeCogs que permite escribir texto en látex, el contexto ontológico generado a partir de la ontología realzada, un cuadro de texto que contiene la cadena de búsqueda, un menú de opciones de búsqueda que permite

seleccionar el tipo de búsqueda a efectuar y un botón de envío que recopilara los datos suministrados para efectuar la operación de búsqueda tal como se muestra en la figura:

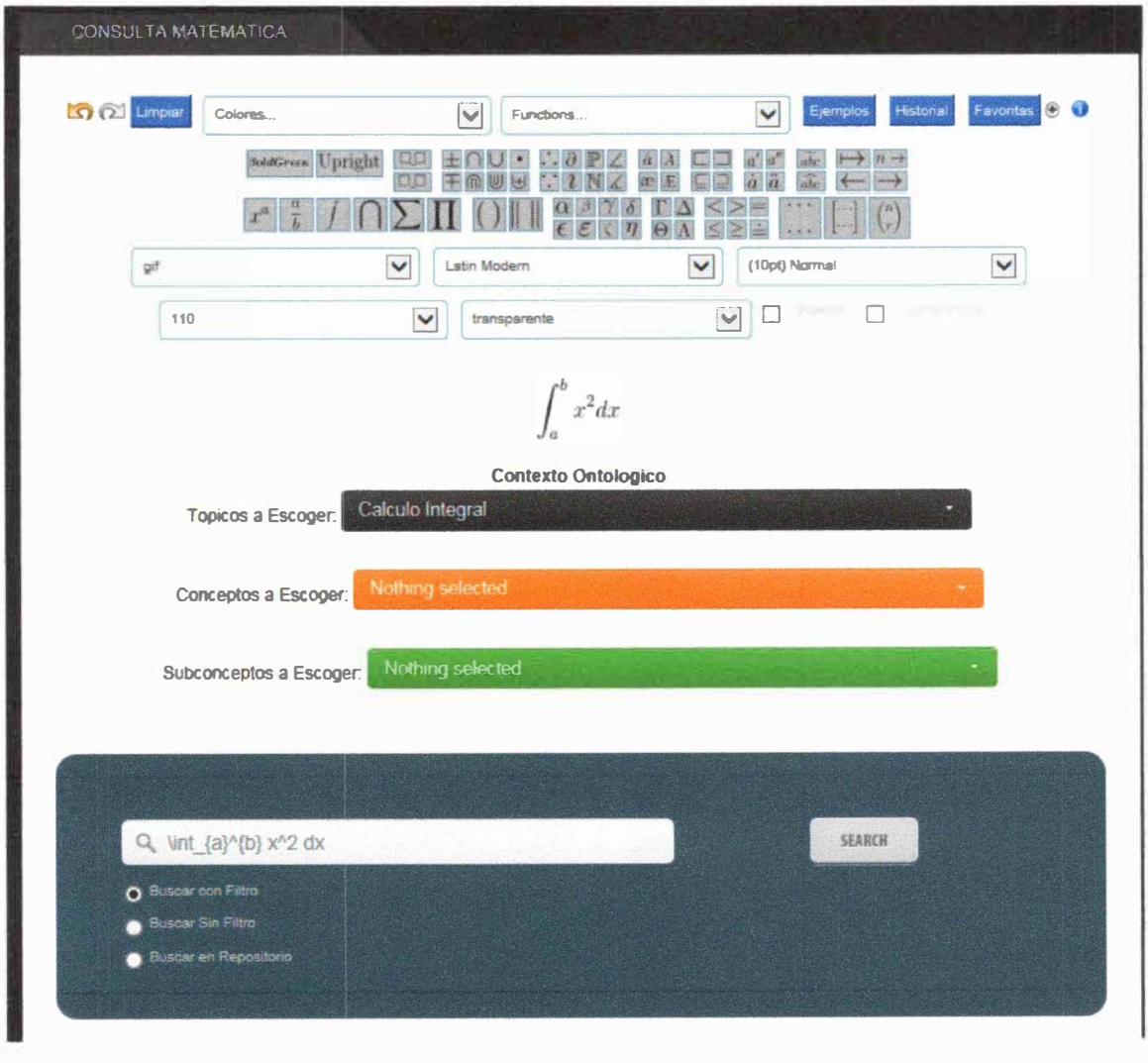


Gráfico 17. Consulta en NOMAT

Fuente: Imagen capturada del Buscador.

Cabe destacar que el editor matemático trabaja en conjunto con el cuadro de texto para ir escribiendo la cadena de consulta, y mediante la selección del contexto ontológico, se determinara en qué tema matemático se clasificarán los resultados encontrados por la herramienta.

4.2.2. Componente de captura de contenidos

En este componente toma la cadena de consulta contenida en cuadro de texto del aplicativo y la envía a un subprograma dentro de la herramienta llamado GoogleApi. GoogleApi trabaja realizando la función de buscar páginas en el repositorio web de google acerca de la cadena en látex que le fue suministrada. Como retorno, la función captura y proporciona un conjunto de páginas web con metadatos (título, url, contenido preliminar y contenido documental de la página entrada) de las cuales serán procesadas en el componente de filtrado.

4.2.3. Componente de Filtrado

El componente de filtrado es uno de los elementos de NOMAT más relevantes puesto en este componente se verifica, clasifica y determina si la página encontrada contiene la cadena solicitada por el usuario.

En primera instancia el contenido de cualquier página web encontrada posee formato HTML, lo que es necesario eliminar para poder acceder a la información contenida en la página web. Para esto NOMAT utiliza un subaplicativo denominado JSOUP el cual es una librería desarrollada en java que permite manipular cualquier texto con dicho formato.

JSOUP posee funciones que permiten eliminar todo tipo de etiquetas HTML que se encuentren en el contenido de las páginas web encontradas por el componente de captura. Utilizando dichas funciones NOMAT tendría acceso a la información general de las páginas ya consultadas.

En segunda instancia se necesita también en eliminar los elementos no significativos en los contenidos de las páginas web tales como los caracteres especiales y las palabras de texto látex que hagan referencia a tipos de letras, colores o espaciado.

Por último, cabe señalar que este componente compara la cadena suministrada junto a cada uno de los contenidos de la página previamente filtrado con el fin de determinar si dicha cadena es contenida en el cuerpo de la página web consultada para así ir creando un grupo de páginas relacionas con la cadena de consulta y agregarlas en última estancia en un arreglo de resultados.

4.2.4. Componente de Recuperación

Este componente se basa en una función de NOMAT que permite almacenar las páginas relacionadas en un arreglo de resultados el cual es enviado y desplegado en el componente de visualización. Cabe destacar que antes de realizar el envío, este componente se encarga de darle un formato HTML para estandarizar los resultados en un paginador que se establecerá en el visualizador de resultados.

4.2.5. Componente de Visualización de la Información

Este componente del sistema toma el arreglo de resultados y lo recorre en cada una de sus posiciones con el fin de obtener los valores e ingresarlos en un paginador de resultados en la misma interfaz del componente de consulta. El paginador muestra las diferentes páginas relacionadas con la búsqueda realizada una tras otra, de forma organizada y con la información referente a la fuente donde se encontró la página web relacionada, el título de la página, contenido preliminar tal como lo muestra la siguiente figura:

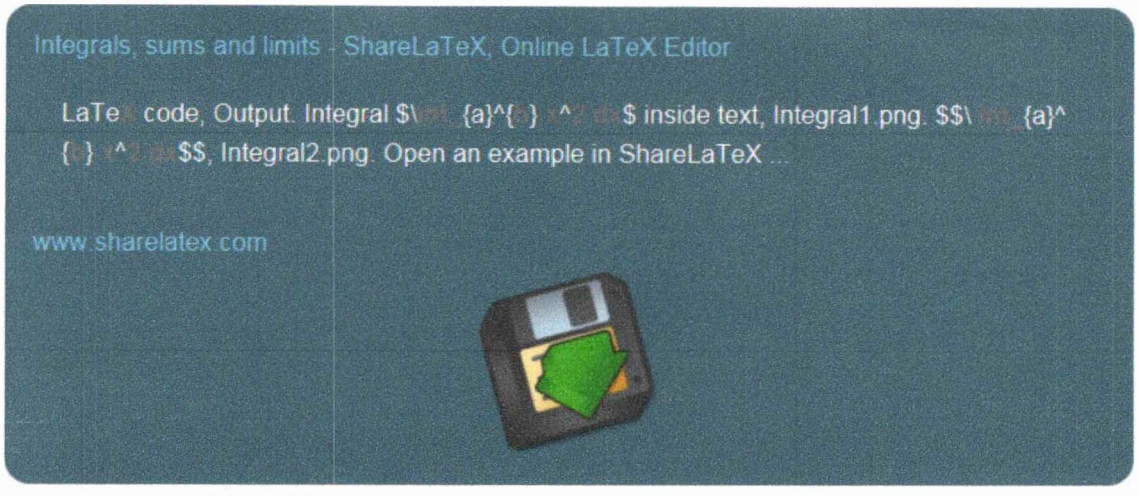


Gráfico 18. Visualización de la Información en NOMAT
Fuente: tomado del aplicativo

El contenido preliminar subraya las ocurrencias de la cadena de consulta destacando en qué lugar y posición de la página que fue encontrada dicha cadena

4.2.6. Componente de Almacenamiento

Es un componente que se encarga de ir almacenando en la base de datos los diferentes resultados obtenidos por las búsquedas realizadas. Cabe destacar que se necesita seleccionar el contexto ontológico primero y luego darle clic en el botón de almacenamiento para poder ir construyendo un repositorio de páginas relacionadas para futuras consultas en NOMAT.

A continuación se muestran las tablas de almacenamiento de los resultados y los contenidos de las páginas relacionadas a búsquedas realizadas en NOMAT

Tabla **asociacion_registros** la cual permite almacenar el criterio de búsqueda

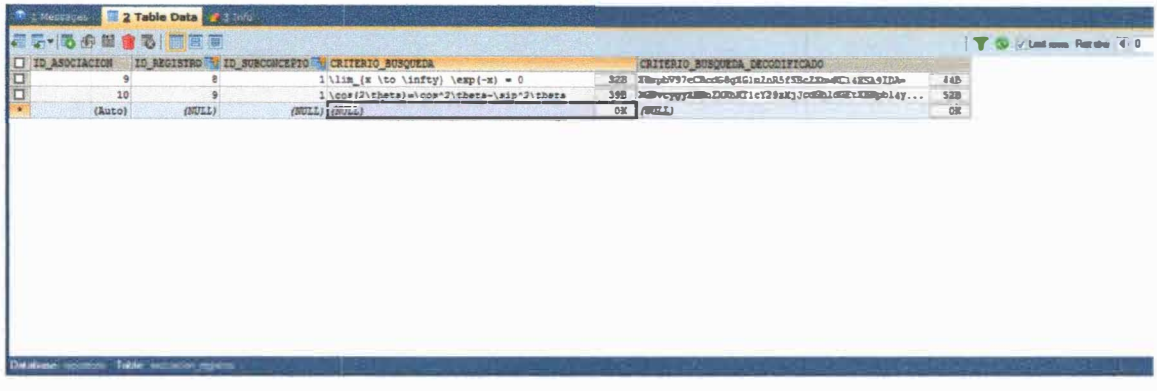


Gráfico 20. Visualización del Criterio de Búsqueda en NOMAT

Fuente: Tomado del aplicativo

Tabla **registros** el cual permite almacenar las páginas resultado obtenidas por los criterios de búsqueda suministrados

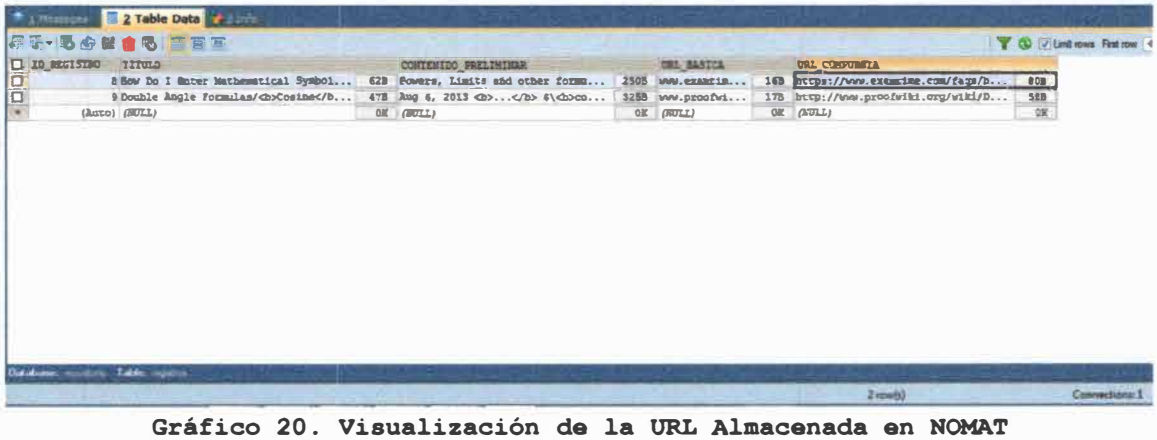


Gráfico 20. Visualización de la URL Almacenada en NOMAT

Fuente: Tomado del aplicativo

4.3. Experimentando con NOMAT

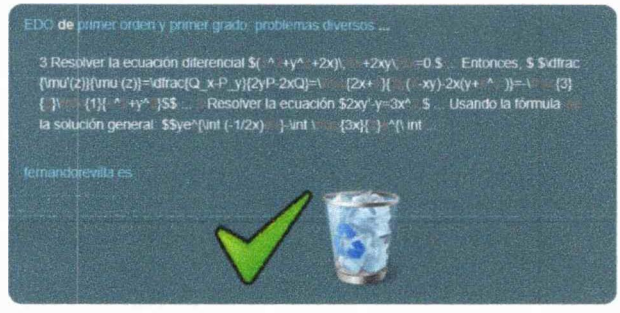
En este apartado se muestra el comportamiento de la herramienta a partir de las mismas consultas hechas en los distintos buscadores. Los resultados se muestran resumidos en la siguiente tabla:

Consulta	Texto en LATEX	Resultados Obtenidos
2x+3y=1	2x+3y=1	Los resultados son muy similares a todos los buscadores, siendo que la consulta se está haciendo en LATEX
Gráfico de 2x+3y=1	Grafico\ de \ 2x+3y=1	Se obtienen páginas relacionadas con gráfico de ecuaciones lineales y en español y el comportamiento es igual a todos los buscadores
$ax^2 + bx + c = 0$	ax^{2}+bx+c=0	Los resultados son muy similares a todos los buscadores
$\frac{d^2y}{dx^2}+5\frac{dy}{dx}+2y=e^x$	\frac{d^2y}{dx^2}+5\frac{dy}{dx}+2y=e^x	Los resultados de consulta muestran ecuaciones diferenciales ordinarias (ver gráfico)
$\int e^{x^2}dx$	\int e^{^{x^{2}}}dx	Los resultados en este caso corresponden solo a búsqueda de expresiones matemáticas que contienen integrales , aunque no son perfectos, está contextualizado (ver gráfico)

Tabla 21. Comportamiento de las Consultas en LATEX hechas por NOMAT

Fuente: Tabla creada por el investigador

UNIVERSIDAD SIMON BOLIVAR
BIBLIOTECA
INSTITUTO DE POSTGRADO



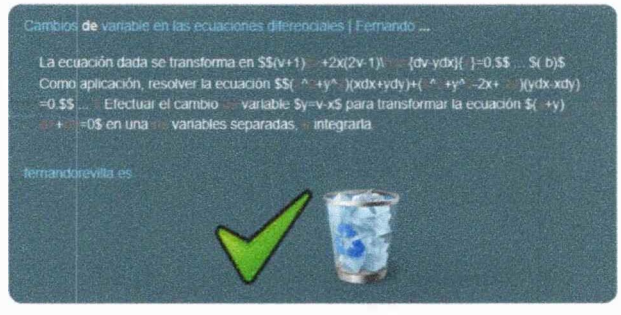


Gráfico 21. Resultados de la consulta de la Ec. Diferencial buscada

Fuente: Gráfico tomado del aplicativo

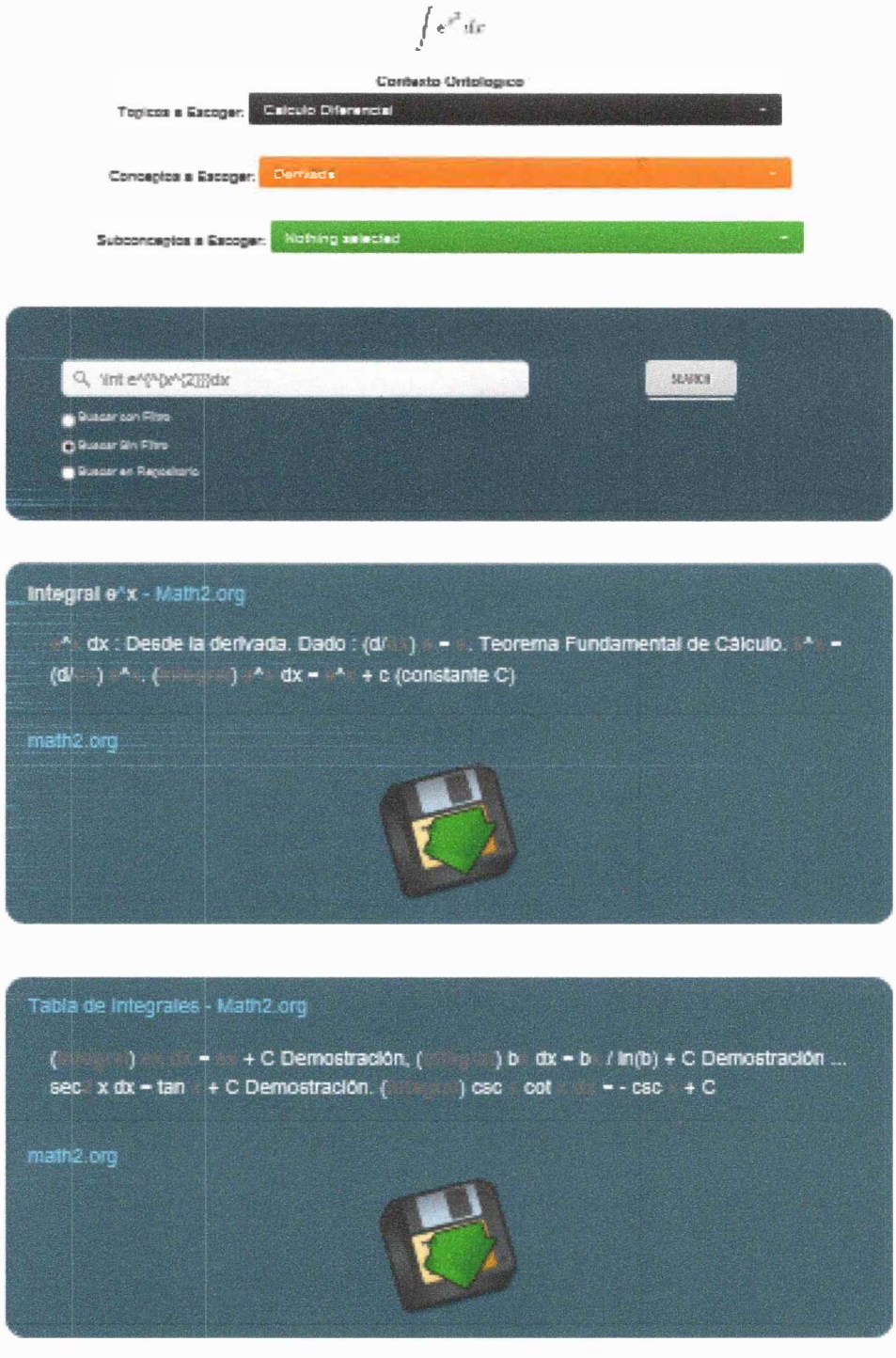


Gráfico 22. Resultados de la consulta de la Integral buscada

Fuente: Gráfico tomado del aplicativo

CONCLUSIONES

El ámbito en el que se enmarca esta tesis de maestría es en el planteamiento de un modelo de buscador de expresiones matemáticas empleando una herramienta ontológica para elaborar el filtrado de la consulta.

La revisión del Estado del Arte realizada evidenció que es poco el material encontrado, sentando esto como precedente que puede ser un nicho de investigación relativamente nuevo en donde ya se han dado algunos pasos.

Luego de un barrido teórico por todo lo relacionado con la Web Semántica y el contexto definitorio hasta llegar al contexto matemático, se puede afirmar que una expresión matemática por sí sola no tiene sentido si no hay un texto complementario que la contextualice. Además, el contexto matemático es un tema muy complicado de abordar por todo lo que ello implica, generando también dificultades para ser implementado en una computadora.

Respecto a la descripción técnica se observa que para el funcionamiento de **NOMAT** se hace necesario el uso de otras aplicaciones convirtiéndolo en una herramienta híbrida. Basado en esto se visualiza la posibilidad de replantear el diseño con el objetivo de sintetizar el uso de estos programas

complementarios cosa que sea mucho más amigable para el usuario final.

Por último, respecto a la experimentación, los resultados obtenidos que fueron sintetizados en la tabla de contrastación muestran que estos son similares, en la medida en que las expresiones sean sencillas y puedan ser transcritas por teclado, son presentadas desde diferentes perspectivas: gráficas, videos, texto, etc., sin embargo al plantear la consulta de ecuaciones diferenciales o integrales o quizá expresiones de mayor complejidad, ninguno de los buscadores comerciales permite la transcripción de las mismas, aspecto que si aborda NOMAT usando a LATEX como transcriptor, arrojando como resultado páginas que han sido editadas con esta herramienta. Demostrando que no solo LATEX no solo puede ser usado como editor de texto científico, sino como editor de texto de consulta, aparte de esto NOMAT se muestra como una herramienta funcional (por utilizar el API de Google y LATEX) y con un cierto grado de efectividad.

6. RECOMENDACIONES Y LINEAS FUTURAS DE INVESTIGACIÓN

Luego de haberse diseñado un prototipo y observar los resultados obtenidos de la experimentación y de contrastar los diferentes resultados, hay algunas recomendaciones a considerar:

En primera instancia no utilizar un API de Google como herramienta de búsqueda sino diseñarlo todo en su totalidad, es decir crear un buscador de forma que se implementen algoritmos de consulta especialmente diseñados para el trabajo con LATEX y analizar la efectividad de estos en la red (y no en repositorios de información).

En torno a los algoritmos de consulta, implementar estos usando tecnología de agentes inteligentes para analizar dicho comportamiento en la red.

Plantear la posibilidad de solicitar la estandarización de las publicaciones de tipo matemático y que estas sean exclusivamente hechas en LATEX ante la W3C que es el ente regulador de las publicaciones en la red a nivel mundial. Esto permitiría establecer consultas de expresiones muy complejas con su respectivo texto complementario, pudiendo no solo hacer

la búsqueda de la imagen de la expresión sino a partir del término escrito editado en LATEX.

Otra alternativa es diseñar herramientas no solo para matemáticos sino para químicos, físicos cuya edición simbólica dependa de LATEX.

REFERENCIAS BIBLIOGRÁFICAS

Web Semántica: Agentes Inteligentes. (2004). Obtenido de http://personales.upv.es/ccarrasc/doc/2003-2004/websemag/agentes.htm#daml

Arquitectura de la Información Web Semántica. (2008). Obtenido de http://luisf.wikispot.org/Web_Semantica

Alarcón, R., & Sierra, G. (2003). El rol de las predicaciones verbales en la extracción automática de conceptos. *Estudios de Lingüística Aplicada*, (págs. 129-144). Barcelona.

Arceo, G., De los Santos Torres, G., & Rivera Reyes, A. (2012). Uso de agentes inteligentes en cajeros automáticos de la Comisión Federal de Electricidad en Villahermosa (México). *XVI Congreso de Ingeniería de Organización*, (págs. 1484-1491). Vigo, España.

Bernaras, A., Laresgoti, I., & Corera, J. (1996). Building and reusing ontologies for electrical network applications. *European conference on Artificial Intelligence* (págs. 298-302). Budapest, Hungría: ECAI_96.

Berners_lee Tim, H. a. (2002). The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities.

Berners-Lee Tim, H. J. (Mayo de 2001). *Scientific American.* Obtenido de Scientific American: http://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf

Berrocal, J., & Figuerola, C. e. (2003). Agentes inteligentes: Recuperacion autónoma de Información en la web. *Revista Española de Documentación Científica Volumen 26 No. 1*, 184-194.

Brenner, W., Zarnekov, R., & Wittig, H. (1998). *Intelligent Software Agents. Foundations and applications.* New York, USA: Springer-Verlag Berlin Heidelberg.

Building and reusing ontologies for electrical network applications, in: Proc. (s.f.). *European Conference on Artificial Intelligence (ECAI-96).* Budapest, Hundría+.

Buitelaar, P. (2009). *Digital Enterprise Research Institute.* Obtenido de Natural Language Processing for the semantic Web: http://www.deri.ie/fileadmin/documents/teaching/tutorials/DERI-Tutorial-NLP.final.pdf

Cabré, T., Estopà, R., & Vivaldi, J. (2001). Automatic term detection. A review of current systems. *Recent Advances in Computational Terminology* (págs. 53-87). Amsterdam, Holanda: John Benjamin Publish.

Carolina García Cataño, D. A. (2002). *Biblioteca Digital y Web Semántica.* Obtenido de http://biblioweb.sindominio.net/telematica/bibdigwebsem.html

Chainbi, W. (2008). An Ontology Based Multi-Agent System Conceptual Model. *In Special Issue of the International Journal of Computer Applications in Technology* (págs. 35-44). Ginebra, Suiza: Inderscience Publishers.

Chen, H., Finin, T., & Joshi, A. (2005). SOUPA: Standard Ontology for Ubiquitous and Pervasive Applications. *Ontologies for Agents: Theory and Experiences* (págs. 233-258). Maryland, USA: Springer.

Contreras, J. (2007). *Tutorial de Ontologías.* Obtenido de http://www.sedic.es/gt_normalizacion_tutorial_ontologias.pdf

Cuenca Grau, B., Horrocks, I., Kazakov, Y., & Sattler, U. (2007). Just the Right Amount:Extracting Modules from Ontologies.

International World Wide Web Conference Committee (págs. 717-726). Banff, Alberta, Canada: WWW 2007.

Cueva, S. P. (2008). *Web Semántica: principios y estándares.* Obtenido de Web Semántica: principios y estándares: http://www.utpl.edu.ec/gcblog/wp-content/uploads/web-semantica.pdf

Cycorp Home of Smarter Solutions. (2013). *OpenCyc.* Obtenido de http://www.cyc.com/platform/opencyc

De Nicola, A., Missikoff, M., & Navigli, R. (2005). A proposal for a Unified Process for ONtology building: UPON. *Database and Expert Systems Applications (DEXA)*, 655-664.

Deagustini, C. A., Falappa, M. A., & Simari, G. R. (2012). *XIV Workshop de Investigadores de Ciencias de la Computación*, (págs. 113-118). Argentina.

Espejo Bohórquez, Y., Téllez, M., & Rodríguez, J. E. (Enero de 2012). *Revista Electrónica Tecnura de la Universidad Francisco José de Caldas, Volumen 16.* Obtenido de Software basado en agentes inteligentes y servicios web para búsqueda de productos en la web: http://www.scielo.org.co/pdf/tecn/v16n31/v16n31a11.pdf

Espino Barrios, L. F. (2008). *Educación basada en la Web Semántica para Universidades Virtuales.* Obtenido de http://www.luisespino.com/pub/educacion_web_semantica_luis_espino.pdf

Fares, R., Costaguta, R., Schiaffino, S., & Missio, D. (2012). *Grupos de aprendizaje colaborativo asistidos por agentes para promover el equilibrio de roles.* Obtenido de WICC Workshop de Investigadores de Ciencias de la Computación, Buenos Aires, Argentina: http://sedici.unlp.edu.ar/bitstream/handle/10915/18386/Documento_completo.pdf?sequence=1

Fernández Poyatos, D., Aguirregoitia Martínez, A., & Boix Martínez, B. (2010). La Web 3.0 y la comunicación turística. *II congreso Internacional de la Comunicación 3.0*, (págs. 49-60). Salamanca.

Fernández, M., Gómez-Pérez, A., & Juristo, N. (1996). METHONTOLOGY: From Ontological Art Towards Ontological Engineering. *Simposio en Ingeniería Ontológica* (págs. 33-40). California, USA: ASSOCIATION FOR THE ADVANCEMENT OF ARTIFICIAL INTELLIGENCE www.aaai.org.

Foundation for Intelligent Pshysical Agents. (2001). *FIPA Device Ontology Specification.* Obtenido de http://www.fipa.org/specs/fipa00091/XC00091c.PDF

García Figuerola Paniagua, C., & Gomez Díaz, R. e. (2009). *Proyecto 7: un motor de recuperación de información web colaborativo.* Obtenido de http://www.ibersid.eu/ojs/index.php/scire/article/viewFile/4016/3650

García Sánchez, F. (2007). *Sistema Basado en Tecnologías del Conocimiento para entorno de servicios Web Semánticos.* Obtenido de http://www.tdx.cat/bitstream/handle/10803/10924/GarciaSanchezFrancisco.pdf;jsessionid=05E5015F00480018E81725385711617C.tdx2?sequence=1

Gerardo Sierra, R. A. (2006). *CORPUS DE CONTEXTOS DEFINITORIOS:UNA HERRAMIENTA PARA LA LEXICOGRAFÍA Y LA TERMINOLOGÍA.* Instituto de Ingeniería, UNAM: México.

Giraldo, J. E., Ruiz, M., & Mateus, S. P. (2011). Modelo para búsqueda y recuperación semántica en bibliotecas digitales. *Revista Facultad de Ingeniería , Volumen 20 No. 30*, 31-40.

González, L., & Echeverri, J. A. (2011). MODELADO CONCEPTUAL DE USUARIOS EN AMBIENTES UBICUOS MEDIANTE AGENTES Y ONTOLOGÍAS.

Revista EIA. Escuela Colombiana de Ingeniería de Antioquia Número 16, 115-126.

Grau, A. (2006). *EOI America*. Obtenido de http://docencia.udea.edu.co/ingenieria/semgestionconocimiento/documentos/Mod7_HerrTec.pdf

Gruber, T. R. (1992). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human and Computer Studies* (págs. 907-928). Minesota USA: Academic Press Inc.

Grüninger, M., & Fox, M. (1995). Methodology for the design and evaluation ontologies. *Workshop on Basic ontological Issues in Knowledge Sharing Montreal* (págs. 109-122). Montreal, Canada: Workshop on Basic ontological Issues in Knowledge Sharing Montreal.

Heckmann, D., Schwartz, T., Brandherm, B., & Kröner, A. (24-29 de Julio 2005). GUMO the General User Model Ontology. *Lecture Notes in computer Science* (págs. 428-432). Alemania: Springer.

Hipola, P., & Vargas Quesada, B. (2000). Agentes inteligentes: definicion y tipologia. Los agentes de informacion. *El profesional de la Información*.

Hofweber, T. (2004). *Stanford Encyclopedia of Philosophy spring 2009 edition*. Obtenido de Stanford Encyclopedia of Philosophy spring 2009 edition: http://plato.stanford.edu/archives/spr2009/entries/logic-ontology/#3.1

Huang, X. (2007). *Semantic Agent Support for Managed Open Information Retrieval Services*. Queen Mary , London.

Intelligent systems research group. (2001). *Arquitectura de Agentes Inteligentes*. Madrid, España: Universidad Politécnica de Madrid.

Juárez Herrero, J. M. (25 de 02 de 2005). *La web semántica. Problemática y arquitectura.* Obtenido de Informas: Revista electrónica de Ingeniería Informática del CIIRM: http://www.cii-murcia.es/informas/abr05/articulos/La_web_semantica_problematica_y_arquiectura.php

Jung, J. J. (2012). Collaborative browsing system based on semantic mashup with open APIs. *Expert Systems with Applications Science Direct Journal.* Korea: Elsevier.

Klusch, M. (1999). *Intelligent Information Agents.* New York: springer verlag.

Kotis, K., & Vouros, G. A. (2006). Human-centered ontology engineering: The HCOME methodology. *Knowledge and information systems No. 10 Vol I*, 109-131.

Leguízamo León, A. V. (2007). *Un acercamiento a la Web Semántica como herramienta para el aprendizaje en línea.* Venezuela.

Lenat D.B., G. R. (1990). *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project.* Boston: Addison-Wesley.

Lozano Tello, A. (2001). *Ontologías en la Web Semántica.* Obtenido de Anobium: http://www.anobium.es/docs/gc_fichas/doc/68ERfhjkmv.pdf

Maracine, V. e. (2012). *Electronic Journal of Knowledge Management Volume 10 Issue 1.* Obtenido de Knowledge use and Sharing into a Medical Community of Practice; the Role of Virtual Agents (Knowbots): https://www.google.com.co/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&sqi=2&ved=0CCwQFjAA&url=http%3A%2F%2Fwww.ejkm.com%2Fissue%2Fdownload.html%3FidArticle%3D320&ei=S10HUbz6CI7k8gSZ0IHQBw&usg=AFQjCNHkFqpJBMtC_KTe0GaeCy6tDHk4zg&bvm=bv.41524429,d.eWU

Márquez Solís, S. (2011). La Web Semántica. En S. Márquez Solís, *La Web Semántica* (pág. 43). España: Santiago Márquez Solís.

Microsoft Corporation. (2012). *Archivos de formatos XML de ejemplo.* Obtenido de MSDN: http://msdn.microsoft.com/es-es/library/ms191234(v=sql.105).aspx

Miguel Zazo Torrubias, V. P. (2011). *SEMSE search: Sistema de recuperación conceptual de esquemas de metadatos.* Madrid, España.

Neches R., F. R. (1991). *Enabling technology for knowledge sharing.* Obtenido de Enabling technology for knowledge sharing: http://tomgruber.org/writing/AIMag12-03-004.pdf

Norvig Peter, R. S. (2003). Artificial Intelligence: A. En R. S. Norvig Peter, *Artificial Intelligence: A.* Upper Saddle River, (ed) Prentice Hall.

Nugent, C., Bridge, D., Murphy, G., & Bernt-Håvard. (2009). *Case-Based Support for Forestry Decisions:How to See the Wood from the Trees.* Obtenido de http://www.cs.ucc.ie/~dgb/papers/Nugent-Et-Al-2009.pdf

Nwana, H. S. (1996). *Knowledge Engineering Review: Software Agents: An Overview.* Obtenido de Intelligent Systems Research Advanced Applications & Technology Department: http://www.sce.carleton.ca/netmanage/docs/AgentsOverview/ao.html

Pardo, L. J. (2006). *Selección de recursos de información disponibles en el Web invisible.* Obtenido de http://brightplanet.com/products/dqm_benefits.asp

Pedro Hipola, B. V.-Q. (1999). Descripción y evaluación de agentes multibuscadores. *El profesional de la Información.*

Peis Redondo, E., & Hassan Montero, Y. (2003). *Ontologías, metadatos y agentes: recuperación "semántica" de la información*. Obtenido de http://libro.ugr.es/bitstream/10481/1206/1/jotri2003.pdf

Poblet, M., Casellas, N., Torralba, S., & Casanovas, P. (2009). Modeling expert knowledge in the mediation domain: a mediation core ontology. *3rd workshop on legal ontologies and artificial intelligence Techiniques (LOAIT'09)* (págs. 109-131). Barcelona, España: Database and Expert Systems Applications (DEXA).

Reuters, T. (2008). *CALAIS*. Obtenido de http://www.opencalais.com/about

Santacruz Valencia, L. P. (2002). *Agentes. Taxonomia y Aplicaciones*. Obtenido de http://www.lite.etsii.urjc.es/liliana/agentes.pdf

Semafora True Semantics. (2012). *Semantic Infraestructure Ontobroker*. Obtenido de http://www.semafora-systems.com/en/products/ontobroker/

Souissi, A., Chainbi, W., & Ghedir, K. (2012). Proposal of New Approach for Ontology Modularization. *6th International Workshop on Modular Ontologies*. Graz, Austria: CEUR Workshop Proceedings.

Sure, Y., & Studer, R. (2002). *On - To - Knowledge Methodology final version*. Alemania: Universidad de Karlsruhe.

Swartout, B., Ramesh, K., Knight, T., & Russ, T. (1997). Toward Distributed use of large scale ontology. *AAAI Symposium on ontological engineering* (págs. 138-149). California, USA: AAAI.

Tim, B.-l. (2000). *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web*. New York: Harper Collins Publishers Inc.

Tim, B.-L. (2000). *Weaving the Web: The Original Design and ultimate Destiny of World Wide Web.* New York, USA: Harper Collins Publishers Inc.

Torcoroma Velásquez Pérez, A. M. (2011). Ontologías: Una técnica de la representación del Conocimiento Volumen 8. *Avances en Sistemas e Informática.*

Uribe Tirado, A. (2004). *Acceso, conocimiento y uso de las herramientas especializadas de internet entre la comunidad académica, científica, profesional y cultural de la Universidad de Antioquia.* Medellín, Colombia.

Uschold, M., King, M., Moralee, S., & Zorgios, Y. (1996). *The enterprise Ontology.* Escocia: Artificial Intelligence Aplications Institute.

Velázquez Muñoz, P. (2008). *Unos métodos en el desarrollo de ontologías.* Obtenido de http://www.gsic.uva.es/wikis/yannis/images/5/5f/TrabajoIndividualFinal_PabloVelazquezMu%C3%B1oz.pdf

Vicente, J., Rebollo, M., & Carlos, C. (2005). Agentes de Información. *INVESTIGACIÓN BIBLIOTECOLÓGICA*, 28-38.

W3C: World Wide Web Consortium. (10 de Febrero de 2004). *OWL Web Ontology Language.* Obtenido de http://www.w3.org/TR/2004/REC-owl-guide-20040210/#OwlVarieties

Wang, X. H., Gu, T., Zhang, D. Q., & Pung, H. K. (Marzo 14 de 2004). Ontology Based Context Modeling and Reasoning using OWL. *Second IEEE Workshop on Context Modeling and Reasoning 2004* (págs. 18-22). Orlando, Florida: IEEE.

Weiss, G. (2001). *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence.* Massachusets, USA: Massachusets Institute of Technology.

Wooldridge, M. (2009). *An introduction to Multiagents systems.* United Kingdom: John Willey & Sons ltd.

W3C, XML Extensible Markup Language (XML) . Recuperado en: http://www.w3.org/XML/

W3C, HyperText Markup Language (HTML) . Recuperado en: http://www.w3.org/community/webed/wiki/HTML

W3C, Cascading Style Sheets (CSS). Recuperado en: http://www.w3c.es/Divulgacion/GuiasBreves/HojasEstilo

W3C, JavaScript Language. Recuperado en: http://www.pergaminovirtual.com.ar/definicion/JavaScript.html

W3C, Jquery JavaScript framework . Recuperado en: http://www.actualidadjquery.es/acerca-de/

W3C, AJAX de Asynchronous JavaScript And XML. Recuperado en: http://www.etnassoft.com/biblioteca/aspects-of-ajax/

Ejemplos de Servlet y JSP Web Application Development Recuperado en: http://webap.wikispaces.com/file/view/JSP_Servlet.pdf

Descripción y características del potente y moderno lenguaje de programación Java, Recuperado en: http://www.desarrolloweb.com/articulos/497.php

Jsoup: Java HTML Analizador Recuperado en: http://jsoup.org/

Editor de ecuaciones latex en línea (CodeCogs) Recuperado en: http://www.codecogs.com/latex/about.php

OWLAPI library http://owlapi.sourceforge.net/

JSON-lib java library, Recuperado en: http://json-lib.sourceforge.net/

Netbeans IDE, Recuperado en: (http://ocw.udl.cat/enginyeria-i-arquitectura/programacio-2/continguts-1/1-introduccioi81n-a-netbeans.pdf)

Protégé, open source ontology editor and knowledge-base framework, Recuperado en http://protege.stanford.edu/

GlassFish servidor de aplicaciones de software, Recuperado en: (http://es.wikipedia.org/wiki/GlassFish)