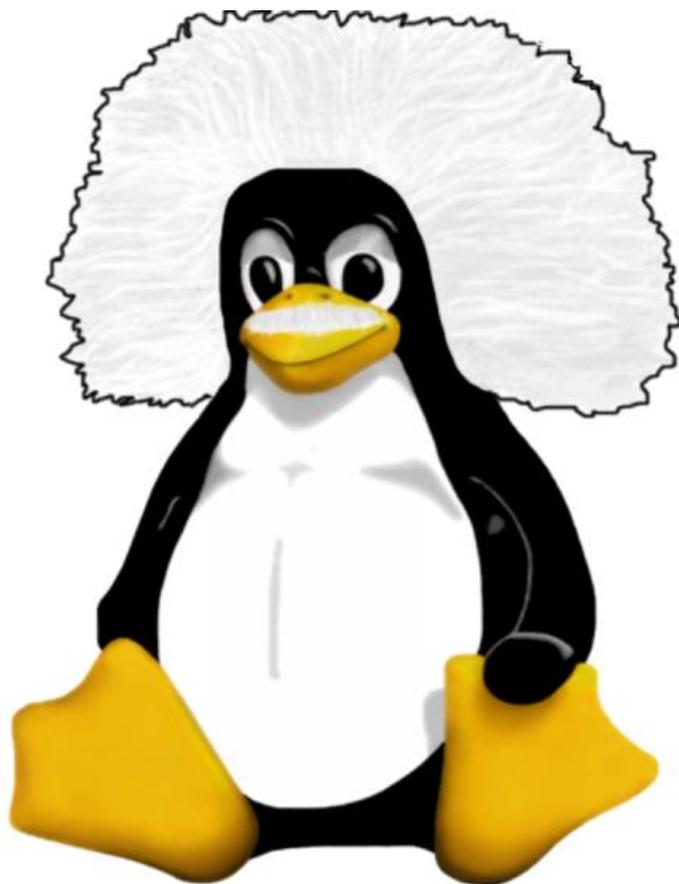


SCIENtific SEarcher: Un sistema para búsqueda y administración de Información Científica



Amanda Kisuy Chi Cerrito
Antonio Téllez Flores
Francisco Federico Raggi Cardenas

Índice general

Portada	1
Contenido	6
I Tutorial	1
1. ¿Qué es ScienSe?	3
2. Antecedentes	5
3. La Historia	7
4. Arquitectura básica del sistema	11
5. Licencia	15
6. Ventajas de ScienSe	17
7. Desarrolladores	21
7.1. Francisco Federico Raggi Cárdenas	21
7.2. Amanda Kisuy Chi Cerrito	22
7.3. Antonio Téllez Flores	23
8. Lo que sigue...	25
II Guía de usuario	27
9. Características de ScienSe	29

10.Requerimientos del lado del cliente	31
11.Agregando publicaciones a la base de datos	33
11.1. Agregando autores	34
11.2. Agregando citas	35
11.3. Consulta de publicaciones	36
11.4. Consulta de autores	37
11.5. Consulta de citas	38
11.6. Consultas en la base temporal de publicaciones	38
11.7. Consultas en la base temporal de autores	40
11.8. Agregando a partir del indexado automático	41
11.9. Comprendiendo los mensajes de error	42
12.Modificando la información de la base	47
13.Obteniendo ayuda y soporte	49
14.Contribuyendo con el proyecto	51
III Guía del administrador	53
15.Requerimientos del sistema	57
15.1. Hardware	58
15.2. Software	58
15.2.1. Para ejecutar ScienSe:	58
15.2.2. Para ver la documentación:	59
15.2.3. Para compilar ScienSe:	59
15.2.4. Para modificar la documentación:	60
15.2.5. Para participar en el proyecto ScienSe:	60
16.Preparando el ambiente	61
16.1. Instalando el software necesario	62
16.2. Configurando PostgreSQL	62
16.3. Configurando apache	63
16.4. Configurando htsciense	64
16.5. Configurando tomcat	64
17.Instalando Science	67

<i>ÍNDICE GENERAL</i>	5
18. Procesos ejecutados por ScienSe	69
19. El proceso de indexado	71
20. El proceso de revisión	73
21. El proceso de añadir información	75
22. El proceso de consulta	77
23. Mantenimiento	79
24. Respaldo de información	81
25. Obteniendo soporte	83
IV Guía del programador	85
26. Estructura básica	87
27. APIs	89
28. Conversores	91
29. Plugins	93
30. Interfaz Web	95
31. La base de datos	97
32. Seguridad	99
V Guía del desarrollador	101
33. Detalles de diseño	105
34. Herramientas de desarrollo	107
35. HTSciense	109

36.La base de datos	111
37.La aplicación web	113
38.Detalles de seguridad	115
39.La documentación	117
40.Futuros Proyectos	119
VI Referencias	123
41.Java Servlets	125
42.Java Server Pages	127
43.Javascript	129
44.Tomcat	131
45.Apache	133
46.Postgresql	135
47.GNU Wget	137
48.RPM	139
VII FAQs	141
49.General	143
50.Usuario	147
Bibliografía	157
Bibliografía	157

Parte I

Tutorial

Capítulo 1

¿Qué es ScienSe?

ScienSe es un sistema de software que pretende proporcionar a investigadores, científicos, estudiantes y en general a cualquier persona que tenga algún tipo de relación con la ciencia, la oportunidad de obtener y manipular de forma rápida, sencilla y eficiente, información relacionada con las publicaciones científicas almacenadas en la Internet o en una intranet local.

ScienSe es un proyecto desarrollado por miembros de la Facultad de Ciencias de la Universidad Nacional Autónoma de México (UNAM) bajo la filosofía del software libre, lo que permite un amplio acceso al sistema y a su mejoramiento y aplicación.

Capítulo 2

Antecedentes

A continuación daremos algunas cuestiones que tienen que ver con la naturaleza de este proyecto y ayudarán a comprender su razón de ser y las motivaciones detrás.

Las publicaciones científicas son importantes por 3 factores fundamentales:

1. Permiten que un investigador reciba críticas a su trabajo de otros investigadores de su misma área, generando un proceso de retroalimentación que mejora el trabajo científico.
2. Permite que nuevos investigadores inicien su trabajo basados en el ya realizado y no comiencen una investigación desde cero.
3. Permite evaluar la calidad de la actividad de los científicos mediante el impacto de sus publicaciones, el cual generalmente se mide por 2 factores; el primero es la publicación en una revista de gran prestigio en el campo y el segundo el número de citas de otros investigadores a su trabajo.

En la actualidad se sufre, por llamarla así, una crisis en las publicaciones de carácter científico, consecuencia de dos fenómenos prácticamente imparables. El primero es el precio abusivo del coste de las suscripciones de las revistas y el segundo la irrupción de Internet.

Los precios de las revistas en los últimos 10 años han elevado sus precios cada año entre el 12 % y el 16 %. Esto ha significado que universidades y bibliotecas de todo el mundo hayan empezado a cancelar títulos de revistas de una forma masiva, pues ya les resulta imposible asumir estos precios. Esto resulta más notable en las universidades de países del tercer mundo donde los gastos en investigación y educación tienden a la baja. El problema es tal, que incluso grandes universidades americanas y europeas han reducido más del 21 % de los gastos previstos para la adquisición de revistas.

El problema ha llegado a algunos científicos y profesores miembros de comités científicos de revistas de prestigio y, sabedores de este desorbitado incremento de precios, han abandonado los propios comités de las revistas ya que no están de acuerdo en volver a pagar por un conocimiento del cual la universidad ya era propietaria y por supuesto no aceptan el absurdo al que se ha llegado, en donde en algunos casos, son los propios investigadores los que deben pagar a las revistas para que sus trabajos sean publicados.

Ante esta situación grupos de investigadores están ya potenciando proyectos alternativos basados en la edición de revistas electrónicas científicas, de acceso libre o bajo coste con acceso vía Internet. Incluso las universidades o grupos de investigadores vuelven a editar sus revistas sin necesidad de intermediarios. Empieza a extenderse la opinión que se debe aprovechar aún más la WWW, nacida en las universidades y centros de investigación, para desarrollar una nueva difusión de la ciencia.

Es en este contexto que se presenta ScienSe, el cual es un Software Libre que pretende contribuir a resolver el problema de difusión de la ciencia, particularmente de las publicaciones de carácter científico. El software libre toma, precisamente la filosofía científica de compartir y ahora la ciencia puede aprovechar este modelo para resolver un problema que preocupa cada día más a las comunidades científicas en todo el mundo.

Aunque han surgido otros proyectos tendientes a resolver este problema, ScienSe combina el software libre con la idea de hacer de libre acceso las publicaciones científicas y basará su éxito en el poder de colaboración de los científicos y los programadores de software libre.

Capítulo 3

La Historia

En Abril de 2003 se llevó a cabo una entrevista entre los pasantes de la licenciatura de Ciencias de la Computación de la Facultad de Ciencias de la UNAM: [Amanda Kisuy Chi Cerrito](#) y [Antonio Téllez Flores](#) con el doctor Francisco Federico Raggi Cárdenas, investigador titular D del [Instituto de Matemáticas](#) de la [UNAM](#), en la que el objetivo principal, era elegir alguno de los problemas a que se enfrentaba el doctor Raggi en su desempeño científico que pudiera ser resuelto por un programa de computadora, el cual sería desarrollado por los alumnos antes mencionados como proyecto de servicio social.

Es así como, explorando las diferentes posibilidades, el doctor Raggi menciona la inconveniencia de, que investigadores como el, no puedan saber quien hace referencia a sus trabajos debido a la gran cantidad de tiempo que se lleva el hacer las búsquedas de esa información manualmente, y que esto origina que en la mayor parte de los casos se pierda un poco las proporciones reales de los trabajos de investigación frente a quienes si se toman el tiempo para realizar tal labor.

Y esa es la idea que da origen a ScienSe.

Con eso en mente Antonio Téllez y Kisuy Chi se enfocan a buscar alternativas de solución.

El primer paso fue realizar cuidadosamente un planteamiento del problema, del cual poder extraer las características fundamentales que un software

debería tener para poder cumplir con ese objetivo. Una vez obtenidos tales requerimientos se realizó una investigación de campo en la que se buscaron y analizaron aquellos sistemas de software que cumplieran con al menos uno de los requerimientos del sistema.

Con esto se pudieron ubicar las desventajas que hacen que dichos sistemas no cumplan con las expectativas de los usuarios. Esto nos llevo a replantear algunos de los requerimientos establecidos al inicio con el fin de no incurrir, en la medida de lo posible, en tales errores.

El siguiente paso fue realizar otra investigación, pero esta vez, enfocada a encontrar la mejor tecnología computacional ya existente que pudiera ser empleada para el desarrollo del sistema.

Debido a que esta búsqueda resultó más fructífera de lo que se esperaba, se noto que las expectativas del sistema podían crecer si se aprovechaban al máximo las características de las herramientas que se emplearían, lo que nos llevó a ampliar nuevamente los requerimientos del sistema.

Después de esto, nos dimos a la tarea de buscar un nombre para el proyecto. Debía reflejar, en la medida de lo posible, la naturaleza del producto y también debía ser capaz de atraer al público al que estaba dirigido. Finalmente, debía ser un nombre fácil de recordar y difícil de olvidar. Fue con estos objetivos que nació la idea de nombrar al producto como SCIENTific SEarcher ([ScienSe](#)), que resultaba fácil de recordar, común, atrayente y lo suficientemente descriptivo.

En esta etapa nos vimos obligados a suspender el proyecto por motivos ajenos a nuestra voluntad.

Fue hasta agosto de 2003 que tuvimos la oportunidad de retomar el proyecto, iniciando con la implementación del software, una vez que hubimos identificado claramente los diferentes módulos del sistema.

Todo marchó como estaba planeado hasta que a mediados de diciembre tuvimos nuevamente que detener el proyecto, debido a una serie de conflictos económicos inesperados. Y el proyecto continuo suspendido hasta febrero de 2004. Aunque esta vez lo reiniciaríamos hasta culminar con la primer versión

en mayo del mismo ao.

Capítulo 4

Arquitectura básica del sistema

Cuando se llevo a cabo la planeación del sistema se pensó en distintas alternativas para diseñar la aplicación, entre ellas se encontraban las siguientes:

1. Un sistema mono-usuario al que se accediera desde una única máquina.
2. Un sistema multi-usuario al que se accede desde una única máquina.
3. Un sistema cliente-servidor donde el cliente es un programa de aplicación.
4. Un sistema cliente-servidor donde el cliente es una interfaz web.
5. Un sistema distribuido donde el cliente puede ser una interfaz web.

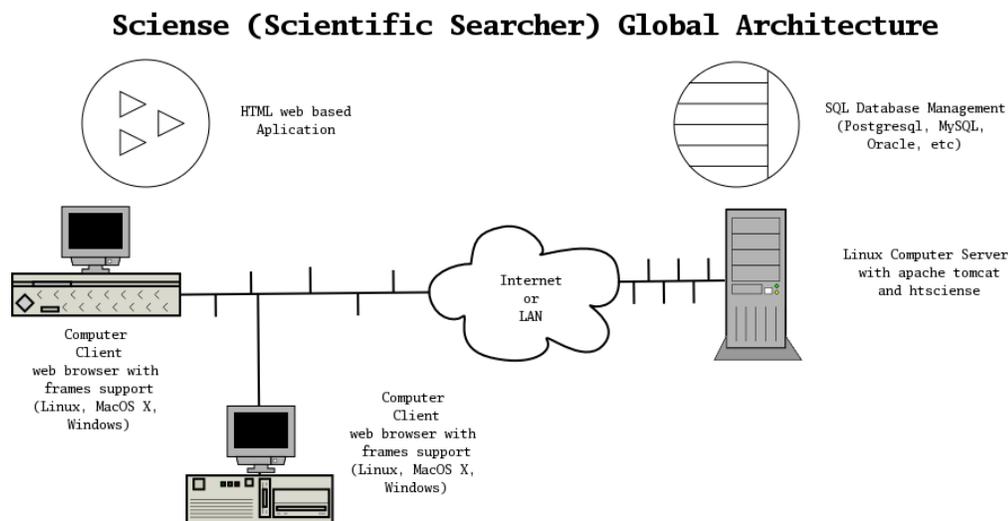
Como la idea principal del sistema es lograr llevar la información científica a la mayor cantidad de personas y tener un sistema muy accesible, se descartaron inmediatamente las opciones 1 y 2.

La opción 3 se tuvo en consideración, pero el esfuerzo de desarrollar clientes para distintas plataformas era demasiado e inútil, incluso utilizando lenguajes de programación altamente portables, pues tendrían que hacerse pruebas de instalación y uso en cada una de ellas, lo cual incrementaría el trabajo y la cantidad de soporte necesario, así que decidimos que lo mejor sería una interfaz web basada en HTML que provee similar funcionalidad sin sus desventajas.

Ya con esto en mente, solo nos quedaban las opciones 4 y 5. De entre estas nos convencía mas la número 5, pues un sistema distribuido haría que disminuyeran los gastos en procesamiento y almacenamiento, además de que mejoraría los esfuerzos de cooperación para compartir información de tipo científico. El problema es que la complejidad del sistema haría que nos tardáramos más en el desarrollo el cual no debería ser muy largo tomando en consideración que solo era un proyecto de servicio social.

Así que al final nos decidimos por la opción 4, aunque teniendo en mente que el sistema podría ampliarse en un futuro a ser un sistema distribuido.

De esta forma ScienSe tiene la siguiente arquitectura básica basada en el modelo cliente-servidor:



El sistema está compuesto por 3 partes principales:

1. **Motor de Indexado**, htscience; basado en htdig (<http://www.htdig.org>), realiza 2 importantes funciones: la primera es indexar la información de los sitios indicados y almacenar esta información en una base de datos para posteriormente realizar búsquedas por medio de un guión CGI; la segunda es que busca artículos de investigación y los analiza para extraer información sobre estos, misma que sirve de entrada a una base de datos relacional.

2. **Base de datos relacional;** Aquí se almacena la información de los artículos de investigación (como son título, autores, ao, citas etc.), en un formato consistente y con una estructura que facilita la búsqueda de información. Esta base se forma a partir de los datos obtenidos automáticamente por el indexador previa revisión de los mismos, aunque también se pueden almacenar datos sin revisión en una base de datos “temporal”.
3. **Aplicación web;** Por medio de un navegador web con soporte de marcos permite consultar toda la información contenida en la base de datos relacional, empleando para ello diferentes criterios de búsqueda, a través de cualquier computadora con acceso al sistema.

Este modelo se eligió porque facilita mucho las tareas de administración, pues al estar todo el software en un sólo punto, las actualizaciones se reflejan inmediatamente en los clientes, aparte de que permite que el sistema pueda funcionar fácilmente en ambientes heterogéneos donde conviven distintas plataformas de hardware/software como sucede en la mayoría de las universidades e institutos de investigación. Otra de las ventajas es que se puede acceder al sistema en forma remota por medio de la Internet o una intranet.

Otro punto importante es que reduce los requerimientos al mínimo, pues un usuario para acceder al sistema únicamente necesita una computadora con un navegador web con soporte de marcos y con acceso por red al sistema.

Capítulo 5

Licencia

ScienSe es un producto de código abierto, protegido por la [Licencia General Pública \(GPL\)](#), lo que permite el uso, modificación y/o revisión de todo el software (incluyendo el código fuente).

La GNU Copyleft, o GPL, permite a los creadores de un programa conservar sus derechos de autor legales, pero permite a los demás copiar, modificar y vender el nuevo programa resultante. Sin embargo, al hacerlo no pueden limitar ningún derecho similar a los que compren el software. Si vende el programa tal como está, o una modificación del mismo, debe facilitar el código fuente.

Para mayor referencia consulte la GPL.

Capítulo 6

Ventajas de ScienSe

A continuación mostraremos algunas de las ventajas que tiene el sistema ScienSe.

Libre

ScienSe esta regido bajo la Licencia General Pública GPL, lo que permite que usted o alguien mas tome, revise, mejore o adapte el código del software sin ningún trámite especial, siempre que no incumpla la GPL a que ScienSe está sujeto.

Gratuito

La mayoría de los sistemas disponibles, requieren del pago de costosas membresias para tener acceso a ellos, lo que origina que el software solo este al alcance de quien quiere y puede pagar por el. En ScienSe pensamos que la información científica debe estar al alcance de quien la requiera en cualquier momento, por lo que ponemos este software a la disposición de quien lo requiera sin esperar algún tipo de remuneración. Aunque por su licencia ScienSe puede ser comercializado, la intención de los autores es contar siempre con una versión del mismo sin mayor costo para el usuario que lo que significa obtener una copia.

Basado en Estándares

Al diseñar ScienSe lo hicimos pensando en la interoperabilidad y por eso lo basamos en estándares que permiten que la información almacenada sea accesible desde cualquier otro programa que maneje los estándares, con lo que el usuario no queda atado al sistema una vez que ha iniciado su uso, como sucede con otros sistemas propietarios; con ScienSe si en algún momento usted decide mudarse de software puede hacerlo sin perder la valiosa información que maneje hasta el momento. Hicimos nuestro mejor esfuerzo para lograr que ScienSe cumpla al menos con los estándares más populares en todos los niveles del sistema.

Automático

A diferencia de otros sistemas existentes ScienSe es mucho más que solo un administrador de información, ya que tiene un indexador integrado. Esto significa que no se requiere que los datos de los artículos sean proporcionados manualmente, ScienSe puede consultar periódicamente los sitios que se le indiquen para obtener de ellos los trabajos que han sido publicados recientemente, obtener los datos relevantes de los mismos, organizar estos datos y almacenarlos en una estructura que facilite su consulta.

Con Tecnología de Avanzada

Utilizamos tecnologías como PostgreSQL, Tomcat, Java Server Pages, Java Servlets, Apache. Además de lenguajes de programación como XML, C++, Java, JavaScript. Todos ellos con amplio soporte y probada calidad, lo que hace que ScienSe garantice un buen desempeño y un fácil mantenimiento. Por ejemplo XML se está convirtiendo en el estándar para el intercambio de información en la web. C++ y Java son de los lenguajes más utilizados tanto en los ambientes académicos como en la industria. Apache, Tomcat y PostgreSQL son aplicaciones con reconocida calidad y algunos de ellos han sido ya utilizados en el medio académico y comercial con gran éxito.

No Elitista

Casi todos los sistemas de este tipo que hay en el mercado, aprovechando la gran necesidad que tienen los usuarios de la información que manejan y la escasez de los mismos, y se dan el lujo de restringir lo mas que pueden el acceso de cualquier forma a sus sistemas.

Son pocos los que pueden pagar la membresia y menos los que tienen el privilegio de gozar de una mención dentro de tales sistemas, principalmente en los países del tercer mundo donde los presupuestos de investigación siguen una tendencia a la baja. Esto origina problemas tales como que si uno tiene el acceso a estos sitios y esta en busca de algún tipo de información no tiene plena certeza de encontrarla en el, debido a que cada uno se restringe a unas cuantos sitios privilegiados, de tal forma que para aumentar de forma significativa esta probabilidad es necesario contar con una membresia en cada uno de los sistemas existentes, cosa que resulta prácticamente ridícula e innecesaria. Con ScienSe este es un problema del pasado ya que debido a que usted mismo puede decidir los sitios donde buscar la información, en cualquier momento puede acceder a ella sin mayores trámites con la seguridad de que usted tiene acceso a la información que ha elegido.

Seguro

El desarrollo del sistema se ha llevado a cabo siguiendo las mejores técnicas para hacer una programación segura en todos los niveles. Además su arquitectura basada en capas hace que la seguridad sea mas fácil de garantizar en cada nivel.

Veraz y flexible

Debido a las características de ScienSe, la información obtenida podrá ser guardada bajo dos modalidades diferentes, dependiendo esto, del proceso que se siga para ello. El proceso de aadir información nueva a la base de datos puede ser uno de los siguientes

- **Automáticamente.** Una vez que el indexador ha terminado su trabajo, uno puede enviar la información obtenida automáticamente a una

base de datos sin necesidad alguna de revisión. Los datos de esta categoría pueden ser consultados en cualquier momento por el usuario cuando este lo solicite, pero con el sobreentendido de que puede existir algún tipo de error en ellos, dado que fueron obtenidos por un algoritmo basado en probabilidades. Si posteriormente el usuario quiere revisar esos datos, podrán entonces ser considerados como parte de los guardados por la categoría siguiente.

- **Confiablemente.** Una vez que el proceso de indexado ha finalizado, el usuario podrá revisar uno a uno los datos obtenidos por el programa permitiendo de esta forma corregir aquellos campos erróneos obtenidos automáticamente y dicha información podrá ser guardada en una base de datos con la información de igual procedencia, con lo que las consultas a esa base de datos ofrecen plena veracidad al usuario.

Eficiente

Debido a la arquitectura de ScienSe cada una de las diferentes capas ha sido implementada de forma lo más eficiente posible, utilizando para ello los mejores algoritmos en la implementación.

Amigable

El equipo de desarrolladores de ScienSe sabe que la facilidad de uso es un factor importante para el éxito de cualquier software. Es por ello que se ha prestado especial cuidado a la interfaz de usuario final. Para ello nos hemos apegado en su diseño a las recomendaciones hechas por los creadores de las interfaces de usuario más exitosas del mercado y a estudios de usabilidad de la literatura.

Capítulo 7

Desarrolladores

Como ya hemos expuesto en la historia del proyecto, el autor de la idea detrás de ScienSe es el Dr. Francisco Raggi, mientras que el diseño y la implementación de ScienSe ha sido efectuada por los prestadores de servicio social Amanda Kisuy Chi Cerrito y Antonio Téllez Flores.

7.1. Francisco Federico Raggi Cárdenas

Actualmente es investigador titular D del Instituto de Matemáticas de la Universidad Nacional Autónoma de México (UNAM). En 1963 obtiene el grado de Matemático, posteriormente realiza los estudios de maestría en la Universidad de Harvard. Y finalmente obtiene el doctorado en la Facultad de Ciencias de la UNAM, siendo la especialidad de este Matemáticas, al presentar la tesis titulada: “Las unidades en anillos de grupo”. Su área de interés actualmente es la teoría de anillos.

Además de su conocida pasión por las matemáticas el Doctor Francisco Raggi tiene muchos otros intereses diversos. Práctica frecuentemente deportes como frontón y tenis. Es aficionado a la lectura y un gran admirador de Jorge Luis Borges. También disfruta ampliamente de las conversaciones sobre política, el cine y la naturaleza.

7.2. Amanda Kisuy Chi Cerrito

Es pasante de la carrera de Ciencias de la Computación de la Facultad de Ciencias de la UNAM. Sus intereses académicos giran en torno a la inteligencia artificial, las ciencias cognitivas, la psicología, el análisis de algoritmos y las redes neuronales. Ha impartido clases de Análisis de Algoritmos, Programación en paralelo, Introducción a Ciencias de la Computación y Programación en C en la facultad de ciencias de la UNAM.

A sus 23 aos Kisuy es poseedora de una gran inteligencia, perseverancia, creatividad y disciplina. A pesar de los problemas y las incertidumbres, pues perdió a su padre y desde pequeña ha tenido que trabajar para ganarse la vida, siempre se ha destacado por ser una de las mejores estudiantes de cada uno de los niveles educativos en que ha estado, obteniendo varias veces reconocimientos a su excelencia académica. Cabe destacar que curso prácticamente toda la carrera de Ciencias de la Computación sin contar con computadora en casa, obteniendo a pesar de ello excelentes calificaciones.

Asimismo ha mostrado un gran interés en la ciencia, participando desde temprana edad en eventos tales como: las olimpiadas de Física, Matemáticas y Química, el concurso Nacional de Matemáticas Pierre Fermat, las Estancias Cortas en el Instituto de Geofísica dentro del programa Jóvenes hacia la Investigación, el Concurso de Matemáticas I, II y III de Iniciación Universitaria, el Congreso Preparatoriano de las Ciencias, entre otras, destacando siempre en cada una de estos.

Se presento como ponente en el “V Coloquio Nacional de Teoría de Códigos, Criptografía y áreas relacionadas”, tras desarrollar junto con Antonio Téllez, un método de cifrado basado en redes neuronales y tiene publicado un artículo de investigación fruto de su estancia en el Instituto de Geofísica de la UNAM.

También gusta de Linux y la programación en C++, C, Java, Prolog, Scheme y Pascal. Es adicta a los juegos matemáticos y de ingenio.

Además de las ciencias sus otros intereses giran en torno a las artes. Obtuvo el primer lugar en el “Concurso Interpreparatoriano de Maquetas de Vivienda Unifamiliar y Multifamiliar” organizado por la Coordinación Gen-

eral de Opciones Técnicas de la Escuela Nacional Preparatoria de la UNAM. Tiene especial gusto por la danza, el patinaje artístico, la opera, la música, el teatro, la arquitectura y la literatura.

También gusta de las actividades deportivas entre ellas, las artes marciales, la gimnasia, el tenis y la natación.

7.3. Antonio Téllez Flores

Pasante de la carrera de Ciencias de la Computación de la Facultad de Ciencias de la UNAM. Interesado en lenguajes de programación, análisis y diseño de algoritmo, sistemas operativos, arquitectura de computadoras y ciencias cognitivas. Ha impartido clases de programación en C, análisis de algoritmos, análisis numérico e introducción a Ciencias de la Computación en la Facultad de Ciencias de la UNAM.

A sus 25 años de edad Antonio es poseedor de una gran imaginación, creatividad, liderazgo y autosuficiencia, que ha sorprendido en varias ocasiones a quienes le conocen por lo innovador de sus ideas, así como por su gran habilidad para la planeación, el diseño y la dirección de proyectos de software. A pesar de su juventud y gracias a sus cualidades científicas, ha participado en congresos exponiendo siempre trabajos de investigación propios.

Es ampliamente reconocido en el área, por sus profundos conocimientos de Linux. Innato programador que maneja lenguajes como: C, C++, Java, Fortran, Haskell, Perl, Scheme, Pascal, entre otros, siendo C++ su lenguaje favorito, hasta el momento.

Como todo buen científico sus conocimientos no se limitan a la computación, pues es un gran adicto a la historia, las matemáticas, la economía, la literatura, la filosofía, la música, el cine, la física, la biología y la psicología.

Su gran pasión por la historia le ha creado gran predilección por grandes historiadores mexicanos como Miguel León Portilla, Elí de Gortari, Jean Meyer, así como por cronistas como: Bernal Díaz del Castillo, Fray Bernardino de Sahagún, Hernando Alvarado Tezozomoc, entre otros: Además de una adicción a los museos y zonas arqueológicas. La economía ha sido para Antonio

una forma científica de entender mejor el mundo.

Su pensamiento se ha encontrado bajo la influencia de autores como: Friedrich Nietzche, Miguel de Cervantes Saavedra, Marques de Sade, Carl Jung, Sigmund Freud, Issac Asimov, Charles Darwin, Desmond Morris, A. Oparin, F. Engels, I. Kant, J. J. Rousseau, V. I. Lenin, Nicolas Maquiavelo, Carl Marx, Rene Descartes, Bertrand Russell, Juan Jose Lopez-ibor, Havelock Ellis, Marvin Minsky etc.

Capítulo 8

Lo que sigue...

En las secciones siguientes presentaremos documentación para aquel que quiera conocer más sobre el proyecto ScienSe, ya sea como usuario final, administrador, programador o como miembro del proyecto, e incluso si simplemente quiere conocer más sobre ScienSe.

Las secciones siguientes son:

- **La guía de usuario:** Contiene información que será de utilidad a los usuarios finales de ScienSe. Habla de los requerimientos y presenta ayuda para utilizar de forma productiva el sistema.
- **Guía del administrador:** Contiene los tópicos que serán de ayuda al administrador de un sitio que tenga instalado o que planea instalar ScienSe, contiene temas como: instalación, configuración y mantenimiento.
- **Guía del programador:** Es una guía que pretende orientar al programador que quiera personalizar ScienSe a sus necesidades particulares.
- **Guía del desarrollador:** Contiene información indispensable para aquel que quiera trabajar en el proyecto ScienSe.
- **Referencias:** Contiene información sobre las tecnologías empleadas en ScienSe.
- **FAQ:** Contiene respuestas de primera mano a preguntas frecuentemente hechas sobre el proyecto ScienSe.

Parte II

Guía de usuario

Capítulo 9

Características de ScienSe

Las principales características de ScienSe son:

1. Indexa documentos desde múltiples sitios web dados por el usuario.
2. Busca documentos en varios formatos incluyendo: texto plano, adobe PDF, Poscript, Microsoft Word, Microsoft Excel y HTML.
3. Organiza y almacena las búsquedas en una base de datos relacional que cumple con el estándar SQL 92.
4. Permite la búsqueda de publicaciones almacenadas en la base de datos utilizando varios criterios, seamos entre ellos ellos: Nombre del Autor, título de la publicación, institución del autor, fecha de publicación, editorial, organización, institución de publicación, tipo de publicación, palabras clave o combinaciones de estas.
5. Posee una interfaz basada en la web sencilla de utilizar.
6. El indexamiento de la información es automático basado en varias heurísticas para buscar las publicaciones y almacenar la información en la base de datos.
7. Es software libre cubierto por la Licencia General Publica GNU.

Capítulo 10

Requerimientos del lado del cliente

Los requerimientos del lado del cliente son mínimos, únicamente es necesario lo siguiente:

1. Un navegador web con soporte de frames como: Netscape, Mozilla, Konqueror, Opera, Galeon, Epiphany, Firefox, Safari, Internet Explorer, etc.
2. Tener acceso por red al sistema: Por ejemplo si el sistema esta disponible en Internet contar con una conexión a Internet, y si esta en una intranet, tener acceso a la misma.
3. También puede ser deseable que su navegador web cuente con soporte de JavaScript y CSS 1.0 o superior, pero no es indispensable, aunque la gran mayoría de los navegadores actuales soporta estas características.

Capítulo 11

Agregando publicaciones a la base de datos

1. Ir a la sección del marco de la izquierda llamada “Add to database”.
2. Seleccionar en el menú desplegable la opción: “Add Publication”.
3. Si en el marco inferior derecho no aparece la página para añadir publicaciones su navegador posiblemente no tenga soporte para Javascript o lo tenga desactivado, en ese caso presione el botón “Go!” que se encuentra abajo del menú desplegable.
4. Una vez que aparece; en el marco inferior derecho, la página titulada “Add Publication” se procede a llenar los datos de la publicación.
5. En caso de que se sienta confundido sobre como llenar algún dato, seleccione con el Mouse el signo de interrogación que este junto al dato y aparecerá una ayuda contextual, ya sea en la forma de una ventana emergente o en la forma de una página web dependiendo del soporte de JavaScript de su navegador. Cabe señalar que cierta información es obligatoria y sin ella el proceso de adición a la base de datos no continuara, esta información obligatoria se encuentra marcada con un asterisco.
6. Una vez que haya terminado de llenar la información de la publicación presione el botón de “submit”, si no hay errores que el sistema pueda detectar como el dejar vacío un campo obligatorio, un ao o un url invalido se presentara otra forma donde de acuerdo al tipo de publicación que haya seleccionado se pedirán datos adicionales.

34CAPÍTULO 11. AGREGANDO PUBLICACIONES A LA BASE DE DATOS

7. Conteste los datos de la misma forma y al final presione de nuevo el botón “submit”.
8. Si su navegador no tiene soporte de JavaScript y la forma ha sido llenada en forma incorrecta se mostrara una página de error que le informara del error cometido.
9. A continuación aparecerá una página que muestra la publicación aadida durante unos segundos y luego se muestra un menú de opciones. Si no quiere esperar puede seleccionar el enlace que dice “Continue”.
10. Si aparece una página con la leyenda “This Publication already exist in the database system” significa que la publicación que quería aadir ya se encontraba en la base de datos.
11. Así finaliza el proceso de aadir una publicación.

11.1. Agregando autores

1. Ir a la sección del marco de la izquierda llamada “Add to database”.
2. Seleccionar en el menú desplegable la opción: “Add Author”.
3. Si en el marco inferior derecho no aparece la página para aadir autores su navegador posiblemente no tenga soporte para Javascript o lo tenga desactivado, en ese caso presione el botón “Go!” que se encuentra abajo del menú desplegable.
4. Una vez que aparece en el marco inferior derecho la página titulada “Add Author” se procede a llenar los datos del autor.
5. En caso de que se sienta confundido sobre como llenar algún dato, seleccione con el Mouse el signo de interrogación que este junto al dato y aparecerá una ayuda contextual, ya sea en la forma de una ventana emergente o en la forma de una página web dependiendo del soporte de JavaScript de su navegador. Cabe sealar que cierta información es obligatoria y sin ella el proceso de adición a la base de datos no continuara, esta información obligatoria se encuentra marcada con un asterisco.

6. Una vez que haya terminado de llenar la información del autor presione el botón de “submit”, si no hay errores que el sistema pueda detectar como el dejar vacío un campo obligatorio, un correo electrónico o un url invalido, a continuación aparecerá una página que muestra los datos del autor aadido durante unos segundos y luego se muestra un menú de opciones. Si no quiere esperar puede seleccionar el enlace que dice “Continue”.
7. Si su navegador no tiene soporte de JavaScript y la forma ha sido llenada en forma incorrecta se mostrara una página de error que le informara del error cometido.
8. Si aparece una página con la leyenda “This Author already exist in the database system” significa que el autor que quería aadir ya se encontraba en la base de datos.
9. Así finaliza el proceso de aadir un autor.

11.2. Agregando citas

1. Ir a la sección del marco de la izquierda llamada “Add to database”.
2. Seleccionar en el menú desplegable la opción: “Add Cite”.
3. Si en el marco inferior derecho no aparece la página para aadir citas su navegador posiblemente no tenga soporte para Javascript o lo tenga desactivado, en ese caso presione el botón “Go!” que se encuentra abajo del menú desplegable.
4. Una vez que aparece en el marco inferior derecho la página titulada “Add Cite” se procede a llenar los datos requeridos en la forma.
5. En caso de que se sienta confundido sobre como llenar algún dato, seleccione con el Mouse el signo de interrogación que este junto al dato y aparecerá una ayuda contextual, ya sea en la forma de una ventana emergente o en la forma de una página web dependiendo del soporte de JavaScript de su navegador. Cabe sealar que cierta información es obligatoria y sin ella el proceso de adición a la base de datos no continuara, esta información obligatoria se encuentra marcada con un asterisco.

6. Se después de de llenar los datos solicitados presione el botón “submit”.
7. Si su navegador no tiene soporte de JavaScript y la forma ha sido llenada en forma incorrecta se mostrara una página de error que le informara del error cometido.
8. Si aparece una página con la leyenda “This Cite already exist in the database system” significa que la cita que quería aadir ya se encontraba en la base de datos. Asimismo también pueden aparecer mensajes de error si alguna de las claves no existe en la base de datos.
9. Así finaliza el proceso de aadir una cita.

11.3. Consulta de publicaciones

1. Ir a la sección del marco de la izquierda llamada “Internal Searching”.
2. Seleccionar en el menú desplegable la opción: “Search Publication”.
3. Si en el marco superior derecho no aparece la página para buscar publicaciones su navegador posiblemente no tenga soporte para Javascript o lo tenga desactivado, en ese caso presione el botón “Go!” que se encuentra abajo del menú desplegable.
4. Seleccione del menú desplegable el tipo de publicación que desea buscar. Por defecto se busca en todos los tipos de publicación.
5. Seleccione el tipo de búsqueda del menú desplegable que desea, por defecto se busca por palabra clave.
6. Escriba el texto a buscar en la casilla de texto, este texto debe de ser de acuerdo al tipo de búsqueda seleccionada. La búsqueda se hace tomando los elementos que comiencen con el texto a buscar. Esta casilla acepta el comodín % que muestra todos los elementos de la base de datos ordenados alfabéticamente dependiendo del tipo de búsqueda.
7. En caso de que se sienta confundido sobre como llenar algún dato, seleccione con el Mouse el signo de interrogación que este junto al dato y aparecerá una ayuda contextual, ya sea en la forma de una ventana emergente o en la forma de una página web dependiendo del soporte de JavaScript de su navegador.

8. Una vez que haya terminado de seleccionar los criterios de búsqueda y haya escrito el texto a buscar, presione el botón “search”. También puede limpiar los campos presionando el botón “clear”.
9. Si no coloco un texto en el campo de texto se mostrara una ventana emergente avisándole o una página de error dependiendo del soporte de JavaScript de su navegador.
10. Si todos el campo de texto fue llenado se mostrara una página con los resultados obtenidos en la búsqueda. Si existen resultados se mostraran listados los primeros 5 resultados. Asimismo al final de la página de resultados se muestra nuevamente el formulario de búsquedas. Cada uno de los campos con los resultados tiene un botón llamado “More Information...” que muestra información detallada de la publicación de acuerdo a su tipo, dentro de esta información también se muestran el número de publicaciones que cita y el número de publicaciones que la citan. Estos resultados se muestran en un enlace que al seleccionarlo nos muestra tales publicaciones.
11. Así finaliza el proceso de buscar una publicación.

11.4. Consulta de autores

1. Ir a la sección del marco de la izquierda llamada “Internal Searching”.
2. Seleccionar en el menú desplegable la opción: “Search Author”.
3. Si en el marco superior derecho no aparece la página para buscar autores su navegador posiblemente no tenga soporte para Javascript o lo tenga desactivado, en ese caso presione el botón “Go!” que se encuentra abajo del menú desplegable.
4. En la página de búsqueda de autores hay 2 campos de texto para buscar por nombre y por institución, al menos uno de los campos deberá ser no vacío y si ambos campos contienen información la búsqueda se hará de tal forma que se cumpla con ambos criterios.
5. En caso de que se sienta confundido sobre como llenar algún dato, seleccione con el Mouse el signo de interrogación que este junto al dato

38CAPÍTULO 11. AGREGANDO PUBLICACIONES A LA BASE DE DATOS

y aparecerá una ayuda contextual, ya sea en la forma de una ventana emergente o en la forma de una página web dependiendo del soporte de JavaScript de su navegador.

6. Una vez que haya terminado de seleccionar los criterios de búsqueda y haya escrito el texto a buscar, presione el botón “search”. También puede limpiar los campos presionando el botón “clear”.
7. Si no coloco un texto en al menos un campo de texto se mostrara una ventana emergente avisándole o una página de error dependiendo del soporte de JavaScript de su navegador.
8. Si al menos un campo fue llenado se mostrara una página con los resultados obtenidos en la búsqueda. Si existen resultados se mostraran listados los primeros 5 resultados. Asimismo al final de la página de resultados se muestra nuevamente el formulario de búsquedas. Cada uno de los campos con los resultados tiene un botón llamado “More Information...” que muestra información detallada del autor, dentro de esta información también se muestran el número de publicaciones, el número de publicaciones que cita y el número de publicaciones que lo citan. Estos resultados se muestran en un enlace que al seleccionarlo nos muestra tales publicaciones.
9. Así finaliza el proceso de buscar un autor.

11.5. Consulta de citas

Para ver como consultar citas véase “Consulta de Autores” y “Consulta de Publicaciones”.

11.6. Consultas en la base temporal de publicaciones

1. Ir a la sección del marco de la izquierda llamada “Automatic Index”.
2. Seleccionar en el menú desplegable la opción: “Search Publication”.

11.6. CONSULTAS EN LA BASE TEMPORAL DE PUBLICACIONES 39

3. Si en el marco superior derecho no aparece la página para buscar publicaciones su navegador posiblemente no tenga soporte para Javascript o lo tenga desactivado, en ese caso presione el botón “Go!” que se encuentra abajo del menú desplegable.
4. Seleccione el tipo de búsqueda del menú desplegable que desea, por defecto se busca por palabra clave.
5. Escriba el texto a buscar en la casilla de texto, este texto debe de ser de acuerdo al tipo de búsqueda seleccionada. La búsqueda se hace tomando los elementos que comiencen con el texto a buscar. Esta casilla acepta el comodín % que muestra todos los elementos de la base de datos ordenados alfabéticamente dependiendo del tipo de búsqueda.
6. En caso de que se sienta confundido sobre como llenar algún dato, seleccione con el Mouse el signo de interrogación que este junto al dato y aparecerá una ayuda contextual, ya sea en la forma de una ventana emergente o en la forma de una página web dependiendo del soporte de JavaScript de su navegador.
7. Una vez que haya terminado de seleccionar los criterios de búsqueda y haya escrito el texto a buscar, presione el botón “search”. También puede limpiar los campos presionando el botón “clear”.
8. Si no coloco un texto en el campo de texto se mostrara una ventana emergente avisándole o una página de error dependiendo del soporte de JavaScript de su navegador.
9. Si todos el campo de texto fue llenado se mostrara una página con los resultados obtenidos en la búsqueda. Si existen resultados se mostraran listados los primeros 5 resultados. Asimismo al final de la página de resultados se muestra nuevamente el formulario de búsquedas. Cada uno de los campos con los resultados tiene un botón llamado “More Information...” que muestra información detallada de la publicación de acuerdo a su tipo. Asimismo se muestra un botón que permite agregar esta publicación en la base permanente.
10. Así finaliza el proceso de buscar una publicación en la base temporal.

11.7. Consultas en la base temporal de autores

1. Ir a la sección del marco de la izquierda llamada “Automatic Index”.
2. Seleccionar en el menú desplegable la opción: “Search Author”.
3. Si en el marco superior derecho no aparece la página para buscar autores su navegador posiblemente no tenga soporte para Javascript o lo tenga desactivado, en ese caso presione el botón “Go!” que se encuentra abajo del menú desplegable.
4. En la página de búsqueda de autores hay 2 campos de texto para buscar por nombre y por institución, al menos uno de los campos deberá ser no vacío y si ambos campos contienen información la búsqueda se hará de tal forma que se cumpla con ambos criterios.
5. En caso de que se sienta confundido sobre como llenar algún dato, seleccione con el Mouse el signo de interrogación que este junto al dato y aparecerá una ayuda contextual, ya sea en la forma de una ventana emergente o en la forma de una página web dependiendo del soporte de JavaScript de su navegador.
6. Una vez que haya terminado de seleccionar los criterios de búsqueda y haya escrito el texto a buscar, presione el botón “search”. También puede limpiar los campos presionando el botón “clear”.
7. Si no coloco un texto en al menos un campo de texto se mostrara una ventana emergente avisándole o una página de error dependiendo del soporte de JavaScript de su navegador.
8. Si al menos un campo fue llenado se mostrara una página con los resultados obtenidos en la búsqueda. Si existen resultados se mostraran listados los primeros 5 resultados. Asimismo al final de la página de resultados se muestra nuevamente el formulario de búsquedas. Cada uno de los campos con los resultados tiene un botón llamado “More Information...” que muestra información detallada del autor, dentro de esta información también se muestran el número de publicaciones. Asimismo se muestra un botón que permite agregar este autor en la base permanente.

9. Así finaliza el proceso de buscar un autor en la base temporal.

11.8. Agregando a partir del indexado automático

1. Ir a la sección del marco de la izquierda llamada “Automatic Index”.
2. Seleccionar en el menú desplegable la opción: “Add Publication”.
3. Si en el marco superior derecho no aparece la página para buscar publicaciones su navegador posiblemente no tenga soporte para Javascript o lo tenga desactivado, en ese caso presione el botón “Go!” que se encuentra abajo del menú desplegable.
4. A continuación aparecerá una página de entrada al sistema que le solicitará un nombre de usuario y una clave. Cabe destacar que en esta área únicamente deberá tener acceso el administrador del sistema. Para saber más sobre quién puede entrar a esta sección véase la “guía de administración” más adelante. En esta página hay 2 campos de texto a llenar los cuales son obligatorios.
5. En caso de que se sienta confundido sobre como llenar algún dato, seleccione con el Mouse el signo de interrogación que este junto al dato y aparecerá una ayuda contextual, ya sea en la forma de una ventana emergente o en la forma de una página web dependiendo del soporte de JavaScript de su navegador.
6. Después de llenar los datos presione el botón “login”. También puede limpiar los campos presionando el botón “clear”. Cabe destacar que en campo de la clave esta no se mostrará por seguridad y únicamente aparecerá un asterisco por cada carácter que usted teclee.
7. Si no coloco texto en ambos campos se mostrara una ventana emergente avisándole o una página de error dependiendo del soporte de JavaScript de su navegador.
8. También si introduce un nombre de usuario y una clave incorrecta se mostrará una página informándole del error y que lo invitara a intentarlo de nuevo.

9. Si introdujo valores correctos aparecerá una página titulada “Add Publication From Automatic Index”. En esta página y en las sucesivas aparece un mensaje que dice: “Click to close this session LogOff” y sirve para invalidar nuestro acceso al sistema y evitar que otro usuario pueda acceder, es muy importante que cuando dejemos de añadir publicaciones o tengamos que dejar nuestra tarea que seleccionemos el enlace marcado en rojo, esto ayudara a la seguridad general del sistema y con ello ayudara a garantizar que la información en el es confiable. En esta página se pregunta si se desea añadir publicaciones del índice automático en forma manual o automática. Si seleccionamos la opción automática todas las publicaciones pasaran directo a la base de datos temporal donde están las publicaciones sin revisión y se nos mostrara una ventana diciéndonos que se completo este proceso con éxito. Por defecto está la opción manual, la cual detallaremos a continuación.
10. Después de seleccionar una opción presione el botón “Yes, Add Publications from the automatic index”.
11. Entonces irán apareciendo una a una las publicaciones a revisar y el usuario podrá editar los campos donde se muestra información, este proceso es similar al de añadir una publicación expuesto secciones atrás. Después de revisar presionamos el botón para agregar a la base de datos y todo el proceso siguiente es idéntico a añadir una publicación.
12. Después de añadir cada publicación se aaden sus referencias si las tiene y el autor de la misma.
13. Al final cuando ya no hay más publicaciones que agregar se le informa al usuario y se le invita a salir de la sesión.
14. Así finaliza el proceso de añadir a la base de datos a partir del indexado automático.

11.9. Comprendiendo los mensajes de error

Existen distintos mensajes de error que pueden aparecer en el uso de ScienSe, en esta sección queremos explicarle el porque ocurren estos errores para que en el futuro evite que aparezcan si esto es posible. Los presentaremos a continuación:

1. “Author Exist In The Database System”: Se quiere aadir a la base un autor que ya esta en la base de datos.
2. ”his Cite already exist in the database system”: Se quiere aadir a la base una cita que está actualmente registrada.
3. “Other user in the system or htsciense obtaining the articles from the web”: Otro usuario con acceso de nivel administrador, esta dentro del sistema o htsciense se encuentra indexando publicaciones. Si intenta acceder al sistema más tarde seguramente el mensaje habrá desaparecido, en caso contrario contáctese con el administrador del sistema o si es usted revise los archivos de log.
4. “The automatic index does not exist”: Esto significa que htsciense no ha iniciado su trabajo o que no ha aadido más publicaciones desde la última vez que usted hizo la revisión. Si persiste este mensaje contáctese con el administrador del sistema.
5. “The Key Of Author is Invalid”: Esto sucede porque en alguna operación que requiere la llave del autor usted introdujo un valor invalido como un campo vacío o una letra en lugar de un número. Generalmente este mensaje solo aparece si su navegador no tiene soporte de JavaScript o lo tiene desactivado, o si quiere enviar una forma inválida de forma deliberada, si es este último caso le invitamos a no hacerlo.
6. “The Key Of Publication Cited is Invalid”: Esto sucede porque en alguna operación que requiere la llave del autor usted introdujo un valor invalido como un campo vacío o una letra en lugar de un número. Generalmente este mensaje solo aparece si su navegador no tiene soporte de JavaScript o lo tiene desactivado, o si quiere enviar una forma inválida de forma deliberada, si es este último caso le invitamos a no hacerlo.
7. “The Key Of Publication is Invalid”: Esto sucede porque en alguna operación que requiere la llave del autor usted introdujo un valor invalido como un campo vacío o una letra en lugar de un número. Generalmente este mensaje solo aparece si su navegador no tiene soporte de JavaScript o lo tiene desactivado, o si quiere enviar una forma inválida de forma deliberada, si es este último caso le invitamos a no hacerlo.

44CAPÍTULO 11. AGREGANDO PUBLICACIONES A LA BASE DE DATOS

8. “The Month Of Publication is Invalid”: Este mensaje solo ocurre si usted quiere enviar una forma inválida de forma deliberada, lo invitamos a no realizar este tipo de acciones.
9. “The name of author is Invalid”: Esto sucede porque en alguna operación que requiere el nombre del autor usted dejó el campo vacío. Generalmente este mensaje solo aparece si su navegador no tiene soporte de JavaScript o lo tiene desactivado, o si quiere enviar una forma inválida de forma deliberada, si es este último caso le invitamos a no hacerlo.
10. “The Title Of Publication is Invalid”: Esto sucede porque en alguna operación que requiere el título de la publicación usted dejó el campo vacío. Generalmente este mensaje solo aparece si su navegador no tiene soporte de JavaScript o lo tiene desactivado, o si quiere enviar una forma inválida de forma deliberada, si es este último caso le invitamos a no hacerlo.
11. “The URL Request is Invalid”: Este caso ocurre cuando usted modifica el url de alguna petición get o trata de realizar una petición get de una página que no soporta este método. Lo invitamos a no hacerlo.
12. “The Year Of Publication is Invalid”: Este error aparece porque en alguna operación que requiere el año usted introduce un año menor a cero o superior al año actual. Generalmente este mensaje solo aparece si su navegador no tiene soporte de JavaScript o lo tiene desactivado, o si quiere enviar una forma inválida de forma deliberada, si es este último caso le invitamos a no hacerlo.
13. “The Key Not exist in the database system”: Es cuando en una operación que requiere de la existencia previa de una llave se da una que no existe en el sistema de base de datos. Asimismo puede aparecer si quiere enviar una forma inválida de forma deliberada, si es este último caso le invitamos a no hacerlo.
14. “Page or resource is not available in this moment”: Este tipo de mensajes aparecen cuando el servidor tiene una gran cantidad de usuarios y no puede atender su petición en este momento preciso. Si este tipo de mensaje aparece muy frecuentemente lo invitamos a que se contacte con el administrador del sistema.

15. “Page or resource not found in this server”: es porque usted solicitó un recurso inexistente en el servidor, lo invitamos a no solicitar páginas que no puedan accederse desde el índice principal.
16. “This Publication already exist in the database system”: Se quiere añadir una publicación que actualmente se encuentra en la base de datos.
17. “Text to Find Missing”: Este tipo de error aparece cuando quiere hacer una búsqueda y deja el campo a buscar vacío. Generalmente este mensaje solo aparece si su navegador no tiene soporte de JavaScript o lo tiene desactivado, o si quiere enviar una forma inválida de forma deliberada, si es este último caso le invitamos a no hacerlo.

Capítulo 12

Modificando la información de la base

En este momento no hay soporte para modificaciones en la base de datos vía la interfaz web. Solo el administrador del sistema puede realizarlo utilizando una operación sql directamente en la consola del sistema de base de datos o utilizando el software SQLTester. Si usted necesita hacer una modificación a los datos contenidos en la base de datos, favor de contactarse con el administrador del sistema, y si usted es administrador del sistema, favor de leer la sección titulada “Guía del Administrador”.

Capítulo 13

Obteniendo ayuda y soporte

Existen distintas formas en las que usted puede obtener ayuda y soporte para ScienSe, las listamos en el orden en que consideramos que usted debe hacerlas para facilitar la labor de encontrar la respuesta.

1. Buscar en la ayuda contextual: Generalmente cada parte de la aplicación cuenta con ayuda contextual, y considerando que hemos trabajado para que la aplicación sea lo más sencilla de utilizar posible, tiene grandes posibilidades de que su duda se solucione revisando la ayuda contextual.
2. Si no pudo encontrar respuesta a su duda puede revisar los FAQ donde es posible que encuentre una solución a su problema o duda.
3. En caso de que no haya encontrado solución todavía puede revisar la sección de ayuda que viene con el sistema para ver si encuentra ahí la respuesta a su interrogante.
4. El siguiente paso a seguir es contactarse con su administrador para ver si el puede ayudarlo.
5. Si lo anterior no es posible, usted es el administrador o no se pudo solucionar su problema visite la página web del proyecto <http://sciense.sourceforge.net> y vaya a la sección de ayuda y soporte, probablemente ahí encuentre solución a sus problemas.
6. Si en ese punto no encontró solución pregunte en los foros o listas de correo de soporte para ver si alguien puede ayudarlo.

7. Si no es así no se desespere, trate de resolver usted mismo el problema, considere que con ScienSe tiene acceso al código fuente para poder solucionar los problemas que aparezcan. Y reporte el problema o error a la lista de errores para poder darle solución lo antes posible, de esta forma estará cooperando con el mejoramiento de este software.

Capítulo 14

Contribuyendo con el proyecto

Existen muchas formas en las que un usuarios puede contribuir con este proyecto. Entre estas se encuentran:

1. Reportando errores encontrados en el software.
2. Reportando errores encontrados en la ayuda.
3. Dando sugerencias sobre como mejorar ScienSe.
4. Participando en los foros de soporte.
5. Implementando mejoras en ScienSe.
6. Escribiendo documentación.
7. Programando plugins.
8. Escribiendo temas para la interfaz web.
9. Ayudando a otros usuarios de ScienSe.
10. Escribiendo paquetes para distintas plataformas.

Para mayor referencia dirigirse a la página del proyecto en la sección de desarrolladores.

Parte III

Guía del administrador

Gracias por utiliza ScienSe (SCIENTific SEarcher), esperamos que le sea de utilidad y le ayude en su labor de investigación. Asimismo lo invitamos a que contribuya con su desarrollo ya sea directamente con código, proponiendo nuevas ideas, enviando reportes de errores o diciéndonos que le gusta y que no le gusta de ScienSe, contribuirá así a hacer de ScienSe un mejor sistema.

Capítulo 15

Requerimientos del sistema

Cuando se desarrollo ScienSe se tuvo en mente que operara con la mínima cantidad de requerimientos tanto de hardware como de software y que los requerimientos de software pudieran ser obtenidos de forma tan sencilla como el propio ScienSe. Asimismo se quería que en cada uno de los requerimientos de software hubiera en el mercado tanto en forma libre como propietaria opciones de donde elegir, aunque por cuestiones de simplificar las pruebas elegimos algunas opciones por defecto.

En esta guía le mostraremos como instalar, configurar y mantener el sistema en optimo funcionamiento. Cada unos de los pasos esta orientado para hacer las cosas lo mas sencillas posible, sin embargo dada la gran cantidad de posibles sistemas y configuraciones siempre pueden ocurrir problemas, si este fuera el caso lo invitamos a revisar el FAQ, solicitar soporte vía correo electrónico en los foros de soporte o si usted mismo soluciono el problema nos escriba contándonos del mismo y como lo soluciono.

Esta guía esta hecha pensando en sistemas con el software por defecto y considerando como sistema operativo un Linux con soporte de RPM. Esto esta motivado por causas de tiempo, ya que fue en plataformas de este tipo en las que hicimos nuestras pruebas. En futuras versiones de esta guía aadiremos ayuda para otras configuraciones, en particular estamos interesados en paquetes para sistemas basados en Slackware, Debian y FreeBSD.

Antes de instalar ScienSe hay que preguntarnos que queremos hacer con ScienSe, pues los requerimientos del sistema variaran ya sea que se quiera

instalar ScienSe en un servidor, participar en el desarrollo, revisar el diseño, adaptar el sistema o construir ScienSe a partir del código fuente.

Una vez identificado que es lo que se quiere hacer con ScienSe se procede a revisar si nuestro sistema cumple con los requerimientos de hardware/software o en su defecto hacemos las acciones que permitan cumplir con estos.

15.1. Hardware

Los requerimientos de hardware dependen en gran medida de los siguientes parámetros:

1. Los componentes de software seleccionados.
2. La cantidad de sitios a ser indexados.
3. La cantidad de posibles usuarios.
4. La cantidad de datos de publicaciones almacenadas.

Y como el sistema ha sido diseñado pensando en la portabilidad, ScienSe puede funcionar prácticamente en cualquier plataforma que pueda ejecutar una implementación de J2EE y cuente con soporte de POSIX.

15.2. Software

A continuación enumeramos el software necesario para las distintas cosas que se quiera hacer con ScienSe. Cabe señalar que primero están las opciones por defecto en las que se ha probado ScienSe y luego se dan alternativas dentro del software comercial o libre/abierto.

15.2.1. Para ejecutar ScienSe:

- **Java Development Kit 1.4 o una versión mas nueva:** Puede utilizarse cualquier entorno de desarrollo Java compatible.
- **Servidor Web Apache 1.3 o una versión mas nueva:** Puede utilizarse cualquier servidor web que soporte CGIs en C.

- **Apache Tomcat 5.0 o versión mas nueva:** Se puede utilizar cualquier servidor compatible con Java 2 Enterprise Edition.
- **Sistema Operativo Linux kernel 2.4 o una version mas nueva:** Se puede utilizar cualquier sistema compatible con POSIX y que cumpla con los demás requerimientos de software como: FreeBSD, OpenBSD, GNU Darwin, MacOS X, Solaris, Windows 2003, etc.
- **PostgreSQL 7.1.2 o una versión mas nueva:** Se puede utilizar cualquier sistema manejador de bases de datos compatible con SQL92 como Firebird, MySQL, Oracle, DB2, etc.
- **Un visor HTML con soporte de marcos:** como Netscape, Mozilla, Konqueror, Opera, Galeon, Epiphany, Firefox, Safari, Internet Explorer, etc. Se recomienda soporte de JavaScript y CSS pero no es indispensable.

15.2.2. Para ver la documentación:

- **Un visor HTML.**
- **Un visor PDF:** Como XPDF, Acrobat Reader, Ghost View, etc.
- **Opcionalmente OpenOffice 1.0 o una versión más reciente:** También se puede utilizar StarOffice 7.0 o superior.

15.2.3. Para compilar ScienSe:

- **Java 2 Software Development Kit versión 1.4 o superior:** Puede utilizarse cualquier entorno de desarrollo Java compatible.
- **GNU Make.**
- **Ant 1.5 o superior.**
- **GNU C++ Compiler 2.95 o versión mas nueva:** Se puede utilizar cualquier compilador ANSI compatible.
- **autoconf.**
- **automake.**

- **Bibliotecas J2EE:** Se pueden usar las de Tomcat.
- **Opcionalmente si se desea hacer paquetes RPM las herramientas para crearlos.**

15.2.4. Para modificar la documentación:

1. **Open Office 1.0 o una versión mas nueva:** También se puede utilizar StarOffice 7.0 o superior.
2. **Dia 0.88 o una versión mas nueva.**
3. **Argo UML.**

15.2.5. Para participar en el proyecto ScienSe:

1. **Los mismos requerimientos que para compilar y para modificar la documentación.**
2. **El sistema de control de versiones CVS.**

Capítulo 16

Preparando el ambiente

Después de analizar los requerimientos tanto de software como de hardware, el siguiente paso es obtener el software necesario. La mayoría de los requerimientos ya se encuentran instalados en su sistema si eligió utilizar la plataforma por defecto. En este apartado únicamente nos referiremos a los componentes que en nuestra opinión son mas difíciles de obtener o no se encuentran disponibles en la mayoría de los sistemas y de nuevo tomando como base los componentes por defecto.

Si usted desea instalar otros componentes y encuentra algún problema, le recomendamos que visite nuestro sitio web en el área de ayuda y soporte, es posible que ahí pueda encontrar ayuda al respecto. También nos gustaría que en caso de que instale el sistema exitosamente en plataformas donde no hayamos realizado pruebas, nos notifique para aadir sus experiencias en la sección de ayuda y soporte.

- J2SE Software Development Kit (SDK) 1.4.2 o superior

Puede obtenerse de la siguiente direccion: <http://java.sun.com/j2se/>.

- Servidor Web Apache 1.3 o superior

Puede obtenerse de <http://httpd.apache.org/> aunque si usa un sistema Linux o FreeBSD le recomendamos que utilice el paquete que viene con su distribución.

- Apache Tomcat 5.0 o superior

Puede obtenerse de <http://jakarta.apache.org/tomcat/> o de <http://www.jpackage.org/> si usa RPMs. Si ya lo tiene instalado en su sistema le recomendamos usarlo y si no le recomendamos, si su sistema soporta RPM, usar el paquete RPM que proporcionamos junto con ScienSe.

- Ant 1.5 o superior

Pueden obtenerse de <http://ant.apache.org/> o de <http://www.jpackage.org/> si usa RPM. igualmente si ya lo tiene instalado en su sistema le recomendamos que lo use, en caso contrario si su sistema soporta RPM le recomendamos utilizar el script `install_ant.sh` que encuentra en la pagina del proyecto que se encarga de obtener vía Internet e instalar ant y todas sus dependencias.

- PostgreSQL 7.3 o superior

Puede obtenerse de <http://www.postgresql.org/>. Pero si ya lo tiene instalado en su sistema le recomendamos usarlo.

16.1. Instalando el software necesario

El procedimiento de instalación de los requerimientos de ScienSe variara dependiendo de su sistema y del paquete, por lo que en el caso de los requerimientos le recomendamos leer la ayuda de cada paquete en particular.

En el caso de ScienSe, este se distribuye en forma de paquete RPM, aunque también puede obtener el código fuente en un archivo comprimido o vía CVS. En el futuro tendremos disponible más tipos de paquetes.

16.2. Configurando PostgreSQL

En en caso de PostgreSQL hay 2 cosas importantes que configura, la primera es crear la base de datos para ScienSe, y la segunda es permitir las conexiones JDBC desde ScienSe a PostgreSQL.

Iniciando PostgreSQL:

1. Si usted ya tiene bases de datos funcionando en su sistema pase a permitiendo conexiones JDBC. En otro caso continúe en el inciso 2.

2. `adduser postgres.`
3. `mkdir /var/lib/pgsql/data`
4. `chown postgres /var/lib/pgsql/data`
5. `su - postgres`
6. `initdb -D /var/lib/pgsql/data`
7. `postmaster -D /var/lib/pgsql/data > logfile 2 > &1 &`
8. `createdb test`

Permitiendo conexiones JDBC

1. En el archivo `postgresql.conf` aadir la línea: `tcpip_socket = true`
2. En el archivo `pg_hba.conf` aadir la línea:

```
host all all ip_server net_mask trust
```

donde `ip_server` es la dirección ip del servidor que tiene a tomcat y `net_mask` es la mascara de red.

16.3. Configurando apache

Por lo general ScienSe trabaja bien con las configuraciones estándar del servidor web apache. La opción más importante para el funcionamiento de ScienSe es que apache permita la ejecución de guiones CGI. El archivo `httpd.conf` deberá tener la siguiente línea:

```
LoadModule cgi_module modules/mod_cgi.so
```

16.4. Configurando htsciense

Htsciense es un software basado en htdig (<http://www.htdig.org>) que lleva a cabo 2 funciones fundamentales:

1. Crea un índice con los sitios indicados y almacena esta información en una base de datos de tipo Berkeley para posteriormente realizar búsquedas por medio de un guión CGI.
2. Busca artículos de investigación y los analiza para extraer información sobre estos, misma que sirve de entrada a una base de datos relacional.

Como se ve, la información más importante que necesita htsciense para funcionar son los sitios que debe revisar, aunque soporta muchas otras configuraciones. En el momento de escribir está ayuda se estaba desarrollando htConf una aplicación Java Swing que facilita la configuración de htsciense por medio de una amigable interfaz gráfica.

La configuración de htsciense se lleva a cabo escribiendo un archivo de configuración htsciense.conf que se ubica en el directorio “/etc” en sistemas Unix. La instalación de ScienSe ya contiene un archivo htsciense.conf por defecto con una configuración que deberá servir para la mayoría de los casos, solo hay que cambiar la opción “start_url” donde deberán ponerse los sitios que se desean revisar.

Se puede ejecutar el comando rundig para iniciar el proceso de indexado, pero no es necesario pues en la instalación se aade una entrada en el comando cron para realizar este proceso todos los días a la media noche. Este valor de cron puede modificarse si así se desea.

16.5. Configurando tomcat

ScienSe funciona prácticamente con la configuración por defecto de tomcat únicamente tiene que aadirse el rol “sciense” y un usuario con ese rol en el archivo de configuración tomcat-users.xml, este usuario será el que tendrá la posibilidad de revisar y aadir las publicaciones obtenidas en forma automática.

Para aadir el rol “sciense” en el archivo tomcat-users.xml (que generalmente se encuentra en /etc/tomcat5) se deberá aadir la siguiente línea:

```
<role rolename="science" />
```

Y para añadir un usuario con este rol en el mismo archivo se deberá añadir una línea como la siguiente:

```
<user username="name" password="pass" roles="science" />
```


Capítulo 17

Instalando Sciense

La instalación depende de si obtuvo sciense en formato war o en formato rpm.

Si la aplicación se obtuvo en formato rpm solo ejecute el siguiente comando.

```
rpm -ihv htsciense-1.0-i386.rpm sciense-1.0-noarch.rpm
```

y la aplicación se encontrara instalada y funcionando, para verla únicamente tiene que abrir la aplicación en tomcat en el path de contexto /sciense.

Si la aplicación fue obtenida en formato war tiene que seguir estos pasos.

1. Abra el tomcat manager en la siguiente dirección.

```
http://localhost:8008/manager/html
```

Nota: El puerto 8008 puede variar dependiendo de la instalación de tomcat.

2. Vaya a la opción de “Archivo war a desplegar”, luego seleccione el archivo sciense.war y presione el botón “desplegar”.
3. La aplicación estará instalada en la trayectoria de contexto /sciense.

Generalmente los valores de configuración por defecto de la aplicación ScienSe funcionarán bien en la mayoría de los casos, aunque pueden ser modificados para adaptarlos a sus necesidades. La mayoría de las configuraciones relacionadas con tomcat se encuentran en los archivos “/sciense/web/WEB-INF/web.xml” y “/sciense/web/META-INF/context.xml”.

El parámetro que se recomienda cambiar es el relacionado con la clave de acceso a la base de datos que se encuentra en el archivo context.xml.

Capítulo 18

Procesos ejecutados por ScienSe

Es importante informar al administrador del sistema sobre los procesos ejecutados por el mismo, para que en caso de que se sospeche de una penetración en el sistema operativo, este consiente de cuales son los procesos normales que deberían de ejecutarse, y más aun tratándose de una aplicación como ScienSe cuya complejidad requiere de la ejecución, a veces en forma simultánea, de una gran cantidad de programas. Los procesos que se ejecutan con ScienSe son los siguientes:

1. **postmaster.** El servidor de postgresql.
2. **httpd.** El servidor web apache.
3. **Java:** El servidor tomcat.
4. **wget:** se encarga de descargar los artículos de la red para ser analizados.
5. **htsearch:** Programa CGI para buscar en el indice.
6. **Htload, htsciense, htdump, htmerge, rundig, htfuzzy, htnotify:** Aplicaciones de la suite de htsciense.

Capítulo 19

El proceso de indexado

Uno de los pasos importantes para poder mantener el sistema funcionando en buen estado es comprender cada uno de los pasos que ejecuta el programa para realizar su trabajo, de entre estos el indexado es uno de los más importantes. El proceso de indexado funciona de la manera siguiente:

En el demonio cron se encuentra registrado el guión rundig, el cual se ejecuta cada semana por las noches para realizar este proceso. rundig ejecuta varios procesos que se encargan de crear el índice de los sitios y analizar las publicaciones que se encuentren, de estos procesos htsciense es el más importante. htsciense entra a los sitios dados por el usuario en el archivo htsciense.conf y comienza el proceso de indexado. Este consiste en crear un índice con las palabras encontradas en el documento, seguir recursivamente los enlaces que encuentre, analizar la información del documento para obtener datos de publicaciones y llenar con esta información un archivo de texto que será utilizado para llenar posteriormente la base de datos.

Cuando htsciense se encuentra con un documento en formato HTML, PDF, PS, DOC, etc, manda a llamar a un proceso externo que ejecuta un filtro para transformar el documento en texto plano y coloca los datos en un archivo temporal del que luego se leen los datos para analizarlos y extraer de ellos la información relevante.

El algoritmo empleado para analizar estos archivos esta hecho con heurísticas basadas en probabilidades.

Mientras htsciense se encuentra realizando el indexado, el archivo de datos donde coloca la información de las publicaciones científicas identificadas se bloquea, con un bloqueo cooperativo para evitar condiciones de competencia con otros procesos de ScienSe.

La información de estas publicaciones se almacena en un formato de tags que puede ser leído luego por una aplicación web.

Capítulo 20

El proceso de revisión

Una vez que htsciense ha llevado a cabo el proceso de indexado, se procede a meter los datos obtenidos a la base de datos relacional. Este proceso puede hacerse con revisión o sin revisión. La revisión tiene que llevarse a cabo, debido a que el algoritmo que identifica los datos de las publicaciones no siempre obtiene datos verdaderos, debido en gran parte a que los artículos en la red se encuentran en una gran variedad de formatos.

Cuando se efectúa la revisión, las publicaciones se almacenan en la base de datos para su posterior consulta, también existe la opción de almacenar los datos sin revisión, esto puede ser requerido, por ejemplo cuando no se tiene el tiempo para hacer esta revisión y quiere obtenerse provecho de la información lo antes posible. Cabe señalar que cuando no se han revisado los datos, estos se almacenan en una base de datos temporal de la que posteriormente, previa revisión, pueden pasar a la base de datos con la información revisada.

Con ello se obtienen ventajas del indexado automático, pero al mismo tiempo se garantiza que la información de la base de datos es confiable.

Capítulo 21

El proceso de añadir información

Aparte de la información obtenida automáticamente, también se puede introducir información por medio de un formulario web. Esta información pasa a la base de datos de publicaciones revisadas. Esta aplicación se conecta vía JDBC a la base de datos para almacenar la información.

Capítulo 22

El proceso de consulta

La consulta de información se lleva a cabo por medio de formularios de una aplicación web. Esta aplicación se conecta vía JDBC a la base de datos para obtener la información requerida.

Capítulo 23

Mantenimiento

El mantenimiento al sistema consiste principalmente en llevar a cabo el proceso de revisión anteriormente descrito y en respaldar la información, proceso que explicamos en la siguiente sección.

Capítulo 24

Respaldo de información

El respaldo de información consiste en respaldar la base de datos y los archivos de configuración. Los archivos de configuración de ScienSe son web.xml, context.xml y htsciense.conf. Para respaldar la base de datos favor de remitirse a los manuales del sistema de bases de datos que utilice, por lo general este será PostgreSQL.

Capítulo 25

Obteniendo soporte

Puede obtener soporte para ScienSe de varias formas, le recomendamos seguir los siguientes pasos, si se encuentra con algún problema.

1. Leer esta guía, es probable que aquí se encuentre la solución a su problema.
2. Revisar el FAQ en la sección de administración.
3. Visite la página web del proyecto <http://sciense.sourceforge.net> y vaya a la sección de ayuda y soporte, probablemente ahí encuentre solución a sus problemas.
4. Si en ese punto no encontró solución pregunte en los foros o listas de correo de soporte para ver si alguien puede ayudarlo.
5. Si no es así no se desespere, trate de resolver usted mismo el problema, considere que con ScienSe tiene acceso al código fuente para poder solucionar los problemas que aparezcan. Y reporte el problema o error a la lista de errores para poder darle solución lo antes posible, de esta forma estará cooperando con el mejoramiento de este software.

Parte IV

Guía del programador

Capítulo 26

Estructura básica

ScienSe es un sistema desarrollado bajo en modelo cliente-servidor. Del lado del cliente tiene una interfaz web basada en HTML, CSS y JavaScript que permite manipular la información contenida en el sistema. HTML es utilizado para organizar la información de menus, formas y tablas de presentación. CSS se encarga de la presentación gráfica al usuario y con JavaScript se validan los datos de entrada y se presentan elementos que facilitan al usuario interactuar con la aplicación.

Del lado del servidor el sistema consta de 3 componentes principales:

El primero de ellos es una base de datos relacional que almacena información sobre publicaciones de carácter científico, la base de datos, y todos los enunciados que manipulan los elementos contenidos en la base de datos se encuentra desarrollados utilizando SQL92 para permitir independencia del sistema manejador de bases de datos.

El segundo componente es una aplicación web del lado del servidor que permite manipular la información contenida en la base de datos. Esta aplicación se encuentra desarrollada utilizando como lenguaje Java y emplea las tecnologías de J2EE, Java Servlets y Java Server Pages. La conexión con la base de datos se realiza por medio de JDBC.

El tercer componente es un programa que genera un índice de una serie de sitios web dados por el usuario, y busca información de publicaciones científicas y almacena estos datos para posteriormente aadirlos a la base

de datos previa revisión por parte del usuario. Fue desarrollado utilizando ANSI/ISO C/C++ y hace uso de las bibliotecas de interfaz con el sistema POSIX.

Capítulo 27

APIs

ScienSe cuenta con una API documentada con el sistema javadoc que permite hacer ampliaciones al sistema para adaptarlo a las necesidades de los usuarios. En un futuro también se contara con APIS para la configuración, para el desarrollo de plugins, y para ampliar htsciense.

Capítulo 28

Conversores

Se pueden desarrollar conversores que conviertan los archivos en distintos formatos a modo texto que pueda ser analizado en forma más sencilla para poder obtener información sobre las publicaciones.

Capítulo 29

Plugins

En la actualidad ScienSe no tiene soporte de plugins pero en futuras versiones se contara con el soporte de plugins para que los programadores desarrollen e integren de forma sencilla y transparente mejoras en los algoritmos de obtención de información de las publicaciones.

Capítulo 30

Interfaz Web

La interfaz web utiliza HTML, CSS y JavaScript para la validación. En la actualidad la interfaz se encuentra en los servlets por lo que para poder modificar esta interfaz es necesario conocer el lenguaje de programación Java y en algunos casos también el funcionamiento de los Java Servlets, en un futuro pretendemos separar la presentación de la lógica permitiendo modificar la interfaz web sin necesidad de conocer estas tecnologías.

Otro punto importante es que la interfaz web debe diseñarse con base en estándares abiertos que permitan que sea visible en la mayor parte de los navegadores web disponibles. Por ahora usamos CSS 1.0, HTML 4.0 y JavaScript 1.5.

Capítulo 31

La base de datos

Se recomienda no modificar la base de datos para evitar romper la integridad de la misma y ofrecer compatibilidad con la aplicación estándar y así poder obtener soporte de forma más sencilla. Dentro de la documentación del programador se encuentran las tablas y los diagramas entidad-relación que permitirán al programador que desee extender a ScienSe diseñar las sentencias SQL.

Cabe destacar que contamos con JDBCTester, una aplicación Java Swing que sirve para probar conexiones JDBC y sentencias SQL. Esta herramienta facilitará al programador su trabajo extendiendo ScienSe o cualquier otra aplicación que utilice JDBC.

Capítulo 32

Seguridad

Al ser ScienSe una aplicación para trabajar en red, la seguridad es uno de las cosas que más hemos tomado en cuenta al diseñarlo y programarlo. Por ello hemos puesto especial énfasis en la validación de datos, en los permisos y en la programación segura.

Parte V

Guía del desarrollador

Esta guía está pensada para orientar al programador que esté interesado en integrarse al equipo de desarrollo o simplemente quiera saber más de las interioridades de ScienSe.

Capítulo 33

Detalles de diseo

ScienSe es un sistema desarrollado bajo en modelo cliente-servidor. Del lado del cliente tiene una interfaz web basada en HTML, CSS y JavaScript que permite manipular la información contenida en el sistema. HTML es utilizado para organizar la información de menus, formas y tablas de presentación. CSS se encarga de la presentación gráfica al usuario y con JavaScript se validan los datos de entrada y se presentan elementos que facilitan al usuario interactuar con la aplicación. Esta validación de los datos se lleva a cabo únicamente para mejorar la interfaz de usuario, pues por motivos de seguridad en los componentes que manipulan la información se vuelve a hacer validaciones. Toda la interfaz se encuentra incorporada en los servlets, por lo que para poder modificarla es necesario conocer esta tecnología.

Del lado del servidor el sistema consta de 3 componentes principales:

El primero de ellos es una base de datos relacional que almacena información sobre publicaciones de carácter científico, la base de datos, y todos los enunciados que manipulan los elementos contenidos en la base de datos se encuentra desarrollados utilizando SQL92 para permitir independencia del sistema manejador de bases de datos. Dentro del código fuente de la aplicación se encuentran las tablas y los diagramas entidad-relación de la base de datos.

El segundo componente es una aplicación web del lado del servidor que permite manipular la información contenida en la base de datos. Esta aplicación se encuentra desarrollada utilizando como lenguaje Java y utiliza las

tecnologías de J2EE, Java Servlets y Java Server Pages. La conexión con la base de datos se realiza por medio de JDBC.

El tercer componente es un programa que genera un índice de una serie de sitios web dados por el usuario y busca información de publicaciones científicas y almacena esta información para posteriormente ser aadida a la base de datos previa revisión por parte del usuario, si encuentra archivos en formatos HTML, PDF, PS, DOC, entre otros, usa un filtro para convertirlos a texto plano y luego aplica el algoritmo de identificación de datos. Toda la información generada es almacenada en una archivo temporal que luego es leída por la aplicación web en el proceso de revisión de publicaciones. Utiliza wget para bajar los documentos y analizarlos. Se encuentra desarrollado utilizando ANSI/ISO C/C++ y hace uso de las bibliotecas de interfaz con el sistema POSIX.

Capítulo 34

Herramientas de desarrollo

En el desarrollo de ScienSe hemos únicamente utilizado herramientas de desarrollo libres, con el fin de hacer accesible a los programadores participar en este proyecto. Entre las herramientas que utilizamos se encuentran autoconf, automake, GNU make, ant, javadoc, Doc++, gdb, CVS, BlueFish, etc. Asimismo se utilizan los lenguajes Java, C/C++, JavaScript, HTML, CSS y SQL92.

Capítulo 35

HTSciense

Este programa esta basado en htdig versión 3.6, en esta versión únicamente se le han hecho modificaciones en la interfaz de usuario y adaptaciones para permitir el proceso de analizar las publicaciones. En un futuro se harán mejoras que tiendan a optimizar sus labores.

Capítulo 36

La base de datos

La base de datos es una base de datos relacional. Ha sido normalizada y tiene mecanismos para verificar la integridad referencial. En todo momento para mantener la portabilidad se han utilizado para su creación sentencias del estándar SQL92.

Capítulo 37

La aplicación web

La aplicación WEB esta desarrollada siguiendo los patrones de diseo de las aplicaciones J2EE. Utiliza JavaServlets y Java Server Pages para la generación automática de contenido, JavaScript para la validación de los datos y para hacer más rica la interfaz de usuario y CSS para mejorar la interfaz.

Asimismo se utiliza una cola de conexiones y manejo de concurrencia en los accesos a la base de datos. Para realizar estas conexiones se emplea JDBC

Capítulo 38

Detalles de seguridad

Aparte de utilizar los mecanismos de seguridad del sistema operativo objetivo y de las aplicaciones que utiliza ScienSe, dentro de la aplicación hemos pensado en la seguridad del sistema.

Siendo los problemas de sobrecarga de buffers, las fugas de memoria y el envío de datos inválidos tres de las principales causas que vulneran la seguridad de los sistemas, hemos puesto especial énfasis en estos puntos.

Con respecto a la sobrecarga de buffers y las fugas de memoria utilizamos Java en la aplicación web, este lenguaje tiene mecanismos de Excepciones y Manejo de memoria que minimizan estos problemas. A su vez, hemos revisado minuciosamente las fugas de memoria en htsciense que esta escrita en C++ y utilizamos la clase string de la Standar Template Library, la cual también contiene mecanismos que ayudan con los problemas de sobrecarga de buffers y fugas de memoria.

Con respecto a el envío de datos inválidos, en la aplicación web, validamos los datos con JavaScript para hacer la interfaz web más rica y también validamos los datos a nivel aplicación, previniendo con ello el envío de datos inválidos por modificaciones en los formularios, o re-escrituras de los urls. En htsciense también verificamos los datos de entrada y tenemos mecanismos que permiten sobreponerse a modificaciones o errores en su entorno de ejecución.

Capítulo 39

La documentación

Es una de las políticas de desarrollo de ScienSe documentar todo el código fuente para facilitar la revisión y modificación del mismo. Esta documentación se debe hacer siguiendo el estándar de documentación de Javadoc tanto para el código en el lenguaje Java como en el lenguaje C++ pues el uso de esta herramienta permite la generación automática de documentación en línea para las diversas APIs.

Además de la documentación del código fuente es importante contar con toda la documentación del proceso de desarrollo utilizando el lenguaje UML Asimismo el diseño de las bases de datos debe documentarse con las tablas y los diagramas entidad-relación para permitir el diseño de consultas con mayor facilidad.

Sobre la documentación de usuario esta debe separarse en dos, por un lado, la documentación que contiene la aplicación no debe contener información demasiado específica que tienda a cambiar con facilidad. Este tipo de información debe aadirse a la sección de ayuda y soporte del sitio web de ScienSe.

Capítulo 40

Futuros Proyectos

Por supuesto el equipo de desarrollo de ScienSe esta consciente de los puntos débiles del producto, por lo que pensando en ello creemos que los siguientes puntos podrían ayudar a la erradicación de tales debilidades.

1. Primero que nada consideramos de vital importancia reforzar el algoritmo de obtención de información de un artículo ya que el que fue implementado en esta versión, al estar basado en probabilidades presenta algunas fallas frente a artículos con formatos fuera de los ordinario. Hemos pensado que una alternativa a este algoritmo es el uso de una red neuronal entrenada debido fundamentalmente, a que estas han demostrado en numerosas ocasiones características ideales para este tipo de trabajo.
2. Otra mejora que podría impactar de forma significativa la usabilidad del sistema es la creación de una interfaz gráfica amigable para la configuración del indexador ya que este proceso por ahora es a través de un archivo de configuración que, aunque fácil de crear y modificar, obliga al usuario final a sumergirse en un mundo de información de cuyo conocimiento podría prescindir, cuando requiere hacer cosas tan simples como añadir una dirección nueva a la búsqueda. La creación de una interfaz que contuviera solo las variables que el usuario debe poder configurar y que fuera capaz de añadir tales valores al archivo de configuración del indexador, permitiría una aun más fácil interacción del usuario con el programa. En la actualidad contamos con una versión de tal aplicación, llamada htConf pero aún esta en fase de desarrollo. También sería deseable permitir la configuración del sistema vía web.

3. Otra de las cosas que están por desarrollarse es JDBCTester una herramienta para probar conexiones JDBC y enunciados SQL, por medio de una interfaz amigable al usuario. En la actualidad ya es funcional pero todavía necesita mejoras en la presentación de los resultados y en el sistema de ayuda.
4. El añadir la posibilidad de hacer búsquedas combinadas es otra de las mejoras que se tienen en cuenta, para esto habría que modificar los componentes que acceden a la base de datos y añadir código a la interfaz para que facilite al usuario el realizar estas búsquedas.
5. Cuando htsciense encuentra archivos del tipo PDF, HTML, DOC, PS, etc. Crea un proceso que manda a llamar un filtro que realiza la conversión del formato dado en texto plano. Estos filtros necesitan mejorarse y añadirse nuevos para manejar otros formatos.
6. En el futuro encontraremos que para distintos tipos de publicaciones unos algoritmos de identificación de datos son mejores que otros, además de que habrá quien quiera probar un nuevo algoritmo sin necesidad de modificar todo el código fuente de htsciense. Por esta razón creemos importante añadir el soporte de plugins para permitir utilizar distintos algoritmos sin necesidad de recompilar la aplicación o conocer los detalles de su funcionamiento.
7. Para mejorar la participación de los usuarios, creemos que es importante que puedan modificar la información de sus publicaciones en todo momento, así como añadir nuevos datos. De tal forma que cada autor participe de manera activa en el proceso de recabar información que haga mas valioso el sistema. Es por ello que un mecanismo como el añadir sesiones de usuarios permitiría mejorar la interfaz y la confiabilidad de ScienSe.
8. Para permitir que ScienSe se adapte mejor a las necesidades de distintos usuarios sin tener que invertir mucho tiempo y recursos en adaptar el sistema a las necesidades particulares queremos introducir en ScienSe el modelo Model-View-Controller en la aplicación, para separa la lógica de la presentación. De esta forma el modificar la interfaz de usuario no requerirá de conocimientos de programación.

9. El uso de marcos fue para mejorar la interfaz de usuario. Pero posteriormente nos dimos cuenta que podría tener problemas en navegadores que no soporten marcos, por esta razón creemos conveniente modificar la aplicación para que no haga uso de marcos.
10. Existen datos que puede ser útil saber sobre una publicación científica y aunque esa información ya puede obtenerse del sistema aún no existe una parte de la interfaz que muestre estos resultados al usuario. Entre los datos que podrían ser de interés se encuentran: los autores más citados, las publicaciones más citadas, etc. Por lo que una posible mejora, sería el añadir una sección que muestra dicha información.
11. Ya con soporte de sesiones sería posible que los usuarios pudieran elegir entre distintas interfaces web, dependiendo de cual sea la que mejor se adapte a sus necesidades, esto por medio de soporte de temas en la interfaz.
12. Dados los pocos recursos con los que realizamos el proyecto no nos fue posible hacer pruebas en distintas plataformas, por lo que sería deseable probar el sistema en distintas plataformas de sistema operativo como MacOS X, FreeBSD, Solaris, Debian, Slackware, por mencionar algunas. Además de probar el sistema en distintas implementaciones de J2EE , en otros sistemas manejadores de bases de datos como MySQL, Oracle, Firebird. Y Elaborar paquetes para las distintas plataformas.
13. Otra de las mejoras que haría de ScienSe una mejor herramienta de trabajo y facilitar el sacarle el máximo provecho sería añadir soporte de JavaHelp, para tener un sistema de ayuda poderoso y consistente.
14. También deseamos mejorar el logo de la aplicación, esperamos que algún usuario con conocimientos de diseño gráfico pueda ayudarnos. Por cierto es requisito es que el logo sea hecho utilizando The Gimp.

Parte VI

Referencias

Capítulo 41

Java Servlets

Los servlets o Java Servlets son una tecnología desarrollada por la compañía Sun Microsystems (<http://www.sun.com>) que proveen un método basado en componentes e independiente de la plataforma para el desarrollo de aplicaciones basadas en la web, sin las limitaciones de los programas CGI. Y sin los mecanismos de extensión del servidor propietarios, los servlets son asimismo independientes del servidor. Esto da la libertad de elegir servidores, plataformas y herramientas de desarrollo.

Capítulo 42

Java Server Pages

Las Java Server Pages (JSP) son una extensión de los Java Servlets, esta tecnología también desarrollada por Sun Microsystems, permite a los programadores y diseñadores web, desarrollar y mantener rápidamente, páginas dinámicas ricas en información. Como parte de la familia de tecnologías Java, las Java Server Pages permiten el desarrollo de aplicaciones basadas en la web que son independientes de la plataforma. La tecnología JSP asimismo, separa la interfaz de usuario de la generación de contenido, permitiendo a los diseñadores cambiar el diseño de la página sin alterar el contenido dinámico subyacente.

Capítulo 43

Javascript

JavaScript es un lenguaje de guiones de Netscape orientado a objetos y de plataforma cruzada. JavaScript es un lenguaje pequeño y ligero; no es útil como un lenguaje independiente, pero está diseñado para ser fácilmente incrustado en otros productos y aplicaciones, tales como navegadores web.

Generalmente es utilizado para proveer interfaces de usuario más ricas y para validar la entrada en formas de aplicaciones web.

Capítulo 44

Tomcat

Tomcat es un contenedor de servlet que es usado como la implementación oficial de referencia para las tecnologías Java Servlet y Java Server Pages. Tomcat es un software abierto distribuido bajo la licencia "Apache Software License".

Capítulo 45

Apache

Apache es el más popular web server en Internet con casi el 70 % del mercado. Implementa los más recientes protocolos web, es altamente configurable y extensible por medio de una API para el desarrollo de módulos, corre en la más amplia gama de plataformas, además de permitir el acceso a todo el código fuente.

Capítulo 46

Postgresql

PostgreSQL es un sistema administrador de bases de datos relacionales pasado en postgres versión 4.2, desarrollado en el departamento de Ciencias Computacionales en la Universidad de California en Berkeley. Postgres fue un pionero en muchos conceptos que aparecieron en algunos sistemas de base de datos comerciales mucho después.

PostgreSQL es un descendiente de fuente abierta del código original de Berkeley. Soporta SQL92 y SQL99 además de otras modernas características, entre ellas:

- Consultas complejas.
- Llaves foráneas.
- Triggers
- Vistas
- Integridad transaccional.
- Control de concurrencia multi-versión.

Asimismo, PostgreSQL puede ser extendido por el usuario en varias formas, por ejemplo añadiendo:

- Tipos de datos
- Funciones

- Operadores
- Funciones de agregacion.
- Lenguajes procedurales.

Y a causa de su licencia, PostgreSQL puede ser usado, modificado y distribuido por cualquiera libre de cargo para cualquier propósito, privado, comercial o académico.

Capítulo 47

GNU Wget

GNU wget es un paquete de software libre para recuperar archivos utilizando los protocolos web más utilizados como son: HTTP, HTTPS y FTP. Es una herramienta de línea de comandos no interactiva, que puede ser llamada desde guiones, tareas cron, terminales sin soporte de X, etc.

wget tiene múltiples características para recuperar archivos grandes o copias de sitios web enteros o sitios FTP fácilmente.

Capítulo 48

RPM

El RedHat Package Manager (RPM) es una poderosa herramienta de línea de comandos desarrollada por la compañía RedHat (<http://www.redhat.com>), para controlar un sistema de administración de paquetes, capaz de instalar, des-instalar, verificar, consultar y actualizar paquetes de software. Cada paquete de software consiste en un conjunto de archivos en un paquete con información acerca del paquete como su versión, descripción, etc. Hay también una API (Application Program Interface), que permite a los programadores avanzados pasar estos comandos y manejar estas transacciones en un lenguaje nativo.

Parte VII

FAQs

Capítulo 49

General

1.1 ¿Qué es ScienSe?

ScienSe es una herramienta de apoyo a la investigación científica. ScienSe busca, organiza y muestra la información primordial de los artículos de investigación y otras publicaciones de carácter científico.

1.2 ¿Qué significa "ScienSe"?

ScienSe es una abreviación de Scientific Searcher que es el nombre largo del sistema.

1.3 ¿De donde surge el nombre?

La elección del nombre fue uno de los procesos más importantes del proyecto debido a las expectativas que se tenían. El nombre debía expresar la esencia del proyecto, además de ser especialmente atractivo al público al que está dirigido. Paso mucho tiempo antes de que se nos ocurriera, pero cuando nos vino a la mente, tuvimos la certeza de que si existía un nombre para este software, debía ser SCIENTific SEarcher.

1.4 ¿Qué licencia tiene ScienSe?

De toda la gama de características atractivas, la licencia es una de las más importantes, debido a que no existe en el mercado ningún otro software con las características de ScienSe y que, como ScienSe, este bajo la licencia general pública (GPL). Esto significa que usted podrá tener acceso al código fuente, y más aún, usted podrá modificarlo para adaptarlo o mejorarlo, de acuerdo a sus necesidades.

1.5 ¿En qué sistemas operativos corre ScienSe?

ScienSe ha sido desarrollado y probado en la plataforma Linux, aunque en principio debería de funcionar sin problemas en cualquier sistema operativo que soporte POSIX y cuente con una implementación de J2EE.

1.6 ¿Dónde obtengo ScienSe?

ScienSe está disponible en varios paquetes en la siguiente dirección en la sección de descargas:

<http://science.sourceforge.net>

Además si deseas probar antes su potencial ya tenemos un sitio corriendo con ScienSe, llamado ScienSeMath que podrás consultar en la siguiente dirección:

<http://itza.fciencias.unam.mx/sciensemh>

Aunque en ScienSeMath solo podrás tener acceso como usuario externo te dará un gran panorama de lo que ScienSe te ofrece. ScienSeMath es un sitio que contiene la información de artículos de investigación de matemáticas de las revistas más relevantes.

1.7 ¿Dónde consigo soporte?

Hemos documentado cuidadosamente cada aspecto relacionado con ScienSe, por lo que hay varias opciones.

- a) La ayuda contextual que se encuentra integrada a la aplicación.
- b) Este FAQ.
- c) Los manuales de usuario, administrador, programador y desarrollador.
- d) En la página de ScienSe en la sección de ayuda y soporte.
- e) En las listas de soporte.

1.8 ¿Qué documentación está disponible?

Los manuales de usuario, administrador, programador, desarrollador, el FAQ, una sección de ayuda y soporte en el sitio de ScienSe, la documentación de la API y la documentación del desarrollo que se encuentra en el código fuente.

1.9 ¿Cómo podría enterarme de los errores conocidos o de las características faltantes?

En mejor lugar para esto es en la sección de ayuda y soporte de la página del proyecto ScienSe.

1.10 ¿Cómo me uno al equipo de desarrollo de ScienSe?

En el área de desarrolladores del sitio principal de ScienSe habrá información al respecto.

1.11 ¿Cómo reporto un error?

En el área de ayuda y soporte hay indicaciones sobre como reportar un error.

1.12 ¿Cómo es ScienSe comparado con el demás software del área?

Tiene muchas de las características de otros sistemas además de gran potencial por su licencia.

1.13 ¿Cómo puedo apoyar económicamente a ScienSe?

Hay información al respecto en la sección Acerca de ScienSe de la página principal del proyecto.

Capítulo 50

Usuario

- 2.1 ¿Qué características debe tener mi navegador para disfrutar de todas las cualidades de la interfaz?

Aunque puedes acceder al sistema con cualquier navegador con soporte de marcos, el sistema funcionará mejor si tu navegador soporta JavaScript y CSS .

- 2.2 ¿Puedo aadir mi información personalmente?

Si puedes, en el manual de usuario te dice como aadir autores, publicaciones y citas.

- 2.3 ¿Puedo obtener una interfaz en mi idioma?

No, y no hay planes para que esto suceda, principalmente porque casi toda la información científica se presenta en idioma ingles y creímos que por esta razón no era necesario ofrecer la interfaz en otros idiomas.

- 2.4 ¿Porqué ScienSe no identifica toda la información de los artículos correctamente?

Existen muchas razones por las que ScienSe no es capaz de identificar toda la información de un artículo correctamente, las más comunes son:

- a) El formato del artículo no es común. ScienSe intenta identificar toda la información de un artículo usando probabilidades, así que, está diseado para identificar la información en los formatos más comunes, y cuando esto no se cumple, ScienSe puede no saber cual es el orden de los campos, así que se confunde. Aunque este no

es un problema que afecte a la mayor cantidad de los artículos, esperamos solucionarlo o mejorar notablemente usando una red neuronal (ver Proyectos Futuros para mayor información).

- b) Los programas que ayudan a convertir el archivo, de cualquiera de los formatos soportados, a texto no pueden convertir el archivo del artículo correctamente. ScienSe puede soportar archivos con diferentes formatos como pdf, ps, rtf y texto, pero para trabajar con un archivo, ScienSe requiere que estén en formato de texto, así que emplea algunos convertidores de formato que convierten esos archivos a texto plano. Desafortunadamente, hay algunos programas que guardan los archivos con algunos caracteres o formatos no estándar, así que los convertidores no pueden obtener toda la información de los archivos con esas características. Así que ScienSe recibe un archivo daado diferente del original.

2.5 He instalado una nueva versión del filtro convertidor, ¿porqué parece que ScienSe ha cambiado? ScienSe puede soportar diferentes formatos, pero para ejecutar el algoritmo de identificación de información, ScienSe requiere que algunos convertidores de formato (como pdftotext) transformen el archivo a texto. Cuando instalas una nueva versión o un convertidor diferente, los archivos de texto plano que recibe para procesar podrían cambiar también, debido a que el programa que los convierte a cambiado, así que el archivo dado a ScienSe podría ser diferente, y esto podría cambiar el archivo dado a ScienSe, por lo que los resultados podrían cambiar.

2.6 ¿Cuál es el formato “ideal”, que ScienSe requiere para obtener la información personal de un artículo?

Como hay muchas formas diferentes de escribir un artículo, no hay forma de estar seguro sobre el orden de la información, así que para mejorar el desempeño de ScienSe, es deseable que los archivos con artículos llenen cierto formato. El formato deseable es:

- a) La primer línea de el archivo debe ser el título del artículo, y debe terminar con un fin de línea (enter).
- b) Los autores deben estar separados pot “,” (comas) excepto por los dos últimos autores que deben estar separados ya sea por la palabra “and” o el caracter &. Por ejemplo:

Raggi Francisco, Urrutia Jorge, Miramontes P., Bracho J. and Viso Elisa Para mencionar: 5 autores.

- c) Si el artículo contiene el mes de publicación la palabra completa en ingles debe estar escrita en lugar de cualquier abreviación. Por ejemplo: January, y no jan, ni enero, o algo por el estilo.
- d) Los emails, si hay alguno, deben ser colocados en el mismo orden en que los autores fueron escritos, y si hay algunos que no tienen dirección de email deberían ser puestos al final. Las direcciones de email pueden ser separadas por ','(comas) o solo con ' ' (espacios). Por ejemplo:

Raggi Francisco, Urrutia Jorge, Miramontes P., Bracho J. and Viso Elisa
 fraggi@matem.unam.mx, urrutia@matem.unam.mx pvm@ciencias.unam.mx

Para:

Raggi Francisco fraggi@matem.unam.mx, Urrutia Jorge urrutia@matem.unam.mx,
 Miramontes P. pvm@ciencias.unam.mx, Bracho J. Viso Elisa

Por simplicidad debe haber a lo más una dirección de email por autor.

- e) Las páginas personales deben estar en el mismo orden en que los autores fueron dados, y pueden estar separados ya sea por un carácter ' ' o ','. Por ejemplo:

Raggi Francisco, Urrutia Jorge, Miramontes P., Bracho J. and Viso Elisa
 fraggi@matem.unam.mx, urrutia@matem.unam.mx pvm@ciencias.unam.mx
<http://www.matem.unam.mx/investigadores/raggi/> <http://www.matem.unam.mx/urrutia/> <http://www.mathmoo.unam.mx/biomat/pedro/pedro.html>,
<http://www.matem.unam.mx/investigadores/roli/>,

- f) El resumen debe iniciar con la palabra “abstract” y debe terminar con un carácter de fin de línea (enter). Por ejemplo:
 .^abstract. Este es un ejemplo.”
 o
 .^bstract
 Este es un ejemplo”

- g)* Los teléfonos debe ser una cadena de números de longitud al menos 5, separados por el caracter '-'. Y deben ser proporcionados en el mismos orden en que los autores. Por ejemplo:

Raggi Francisco, Urrutia Jorge, Miramontes P., Bracho J. and Viso Elisa 50-00-00-00 51-11-11-11 01-23-45-56-78-9 98-76-54

Para:

Raggi Francisco 50-00-00-00,
 Urrutia Jorge 51-11-11-11,
 Miramontes P. 01-23-45-56-78-9,
 Bracho J. 98-76-54,
 Viso Elisa ,

- h)* El nombre de la revista o diario debe ser dado usando cualquiera de las siguientes palabras, y debe terminar con un fin de línea (enter):
- 1) "journal". Ejemplo: Linux Journal
 - 2) "magazine". Ejemplo: Science Magazine
 - 3) "communications". Ejemplo: ACM Comunications
 - 4) "scientific". Ejemplo: Scientific American
 - 5) "proceeding". Ejemplo: Proceedings of the National Academy of Sciences
 - 6) "book". Ejemplo: Algebra's book
 - 7) "review". Ejemplo: ACM Operating Systems Review
 - 8) "nature". Ejemplo: nature.com
 - 9) "science". Ejemplo: Science Online
- i)* La seccion de las Referencias debe ser la última sección del artículo, y debe comenzar con la palabra "references" (sin importar si esta en mayúsculas o minúsculas).
- j)* Cada referencia debe comenzar de una de las siguientes formas:
- 1.- [algo] algo
 - 2.- (algo) algo

3.- algo algo

4.- algo.- algo

Donde algo significa cualquier caracter, número o letra diferente de los que contiene el formato (1, 2, 3 ó 4) que se este usando.

NOTA: La primer referencía que se encuentre mostrará la forma en que el resto de las referencias van a ser esperadas, así que si la primera comienza con [1] las siguientes deben ser dadas con [y].

Por ejemplo:

[1] Marvin Minsky, "The Society of Mind", Simon and Shuter, 1985, 300p. [2] Marvin L. Minsky and Seymour A. Papert, Perceptrons: Introduction to Computational Geometry, (1988), MIT Press

o

(1) Marvin Minsky, "The Society of Mind", Simon and Shuter, 1985, 300p.

(MM) Marvin L. Minsky and Seymour A. Papert, Perceptrons: Introduction to Computational Geometry, (1988), MIT Press, 34-40.

o

1 Marvin Minsky, "The Society of Mind", Simon and Shuter, 1985, 300p.

a Marvin L. Minsky and Seymour A. Papert, Perceptrons: Introduction to Computational Geometry, (1988), MIT Press, 34-40.

o

SM.- Marvin Minsky, "The Society of Mind", Simon and Shuter, 1985, 300p.

1.- Marvin L. Minsky and Seymour A. Papert, Perceptrons: Introduction to Computational Geometry, (1988), MIT Press, 34-40

2.7 .- ¿Cuál es formato deseable para dar las referencias de un artículo de acuerdo con ScienSe?

Cada cita debe ser dada con el siguiente formato:

- a) Las páginas del diario donde está el artículo, debe ser dada ya sea por dos números separados por un '-' o por un número que termine con "p". Ejemplo:

[SM] Marvin Minsky, "The Society of Mind", Simon and Shuter, 1985, 300p. [P] Marvin L. Minsky and Seymour A. Papert, Perceptrons: Introduction to Computational Geometry, (1988), MIT Press, 34-40

- b) El ao de publicación debe ser mayores a 1900 y menores al presente ao (de la computadora donde ScienSe esta corriendo). El ao puede ser dado como el número solo o como (número) con los caracteres '('y ')'.
Ejemplo:

Marvin Minsky, "The Society of Mind", Simon and Shuter, 1985

Para:

ao=1985

o

Marvin L. Minsky and Seymour A. Papert,
Perceptrons: Introduction to Computational Geometry, (1988),
MIT Press

Para:

ao=1988

- c) El diario debe ser la última cadena de la referencia despues de quitar el ao y las páginas. El diario es considerado como toda la cadena antes de la última ','(coma), así que el diario no debe contener algún ',' dentro de la cadena.

Ejemplo:

Marvin L. Minsky and Seymour A. Papert,
Perceptrons: Introduction to Computational Geometry, MIT Press

Para:

diario = "MIT Press"

- d) El título debe ser la última cadena de la referencia después de borrar el ao, las páginas y el diario. Se considera que el título es toda la cadena antes de la última ',' (coma), así que el título no debe contener alguna ',' dentro de la cadena.

Ejemplo:

Marvin L. Minsky and Seymour A. Papert,
Perceptrons: Introduction to Computational Geometry

Para:

título=Perceptrons: Introduction to Computational Geometry

- e) Los autores son considerados como contenidos en la cadena de la referencia después de borrarle el año, las páginas, el diario y el título. Cada autor es separado de los demás, por el carácter ',' y puede contener la palabra "and" o el carácter '&' separando los dos últimos autores. NOTA: como se supone que cada autor está separado de los otros por una ',' ningún nombre de autor debe contener ese carácter. Ejemplo:

Marvin L. Minsky and Seymour A. Papert

o

Raggi Francisco, Urrutia Jorge, Miramontes P., Bracho J. and
Viso Elisa

o

Raggi Francisco, Urrutia Jorge, Miramontes P., Bracho J., Viso
Elisa

2.8 ¿Es el formato ideal el único formato con el que ScienSe puede trabajar?

No. ScienSe trabaja razonablemente bien (obtiene arriba del 50 % de la información correctamente) con casi cualquier formato, pero las probabilidades de error incrementan cuando los artículos no tienen un formato común.

Bibliografía

- [1] JavaServer Pages Specification; version 2.0; Proposed Final Draft 3. Copyright 2003 Sun Microsystems, Inc.; April 11, 2003.
- [2] Unified Modeling Language (UML) Specification: Infraestructure; version 2.0.
- [3] A Simplified Introduction to \LaTeX ; Harvey J. Greenberg; University of Colorado at Denver; April 18, 2000.
- [4] Maximum RPM; edition 2; Edward C. Bailey; Red Hat, Inc.; 2000.
- [5] Designing Enterprise Applications with the J2EE Platform; Second Edition; Inderjeet Singh, Beth Stearns, Mark Johnson, and the Enterprise Team; Addison-Wesley.
- [6] PostgreSQL 7.3.2 Administrators Guide; PostgreSQL Global development Group.
- [7] PostgreSQL 7.3.2 Programmers Guide; PostgreSQL Global development Group.
- [8] PostgreSQL 7.3.2 Users Guide; PostgreSQL Global development Group.
- [9] PostgreSQL 7.3.2 Reference Manual; PostgreSQL Global development Group.
- [10] Web Style Guide: Basic Design Principles for Creating Web Sites; 2nd Edition; Patrick J. Lynch, Sarah Horton; www.webstyleguide.com.
- [11] GNOME Human Interface Guidelines; The GNOME Usability Project; Calum Benson, adam Elman, Seth Nickell, Colin z Robertson.

- [12] Java Look And Feel Desing Guidelines; Second Edition; Sun Microsystems, Inc.
- [13] KDE User Interface Guidelines.
- [14] The J2EE 1.4 Tutorial; Eric Armstrong, Jennifer Ball, Stephanie Bodoff, Debbie Bode Carson, Ian Evans, Dale Green, Kim Haase, Eric Jendrock; March 31, 2004; Sun Microsystems, Inc.
- [15] Cascading Style Sheets, level 1; W3C Recommendation 17 Dec 1996, revised 11 Jan 1999; Hakon Wium Lie, Bert Boos; W3C.
- [16] HTML 4.01 Specification; W3C Recommendation 24 December 1999; Dave Raggett, Arnaud Le Hors, Lan Jacobs; W3C.
- [17] Web Server Administrators Guide to the Robots Exclusin Protocol.
<http://www.robotstxt.org>
- [18] Macintosh human Interface Guidelines; Apple Computer, Inc.; Addison-Wesley Publishing Company.