



Comisión Nacional de Investigación
Científica y Tecnológica - CONICYT



COMISIÓN NACIONAL DE INVESTIGACION CIENTÍFICA Y TECNOLÓGICA

VERSION OFICIAL Nº 2

FECHA: 04/10/2011

Nº PROYECTO : 1080672 **DURACIÓN :** 2 años **AÑO ETAPA :** 2009

TÍTULO PROYECTO : DATA MINING FOR CONTENT SYNDICATION IN THE WEB

DISCIPLINA PRINCIPAL : INGENIERIA EN COMPUTACION

GRUPO DE ESTUDIO : INGENIERIA 2

INVESTIGADOR(A) RESPONSABLE : CARLOS ALBERTO HURTADO LARRAIN

DIRECCIÓN :

COMUNA :

CIUDAD :

REGIÓN : METROPOLITANA

FONDO NACIONAL DE DESARROLLO CIENTIFICO Y TECNOLOGICO (FONDECYT)

Moneda 1375, Santiago de Chile - casilla 297-V, Santiago 21

Telefono: 2435 4350 FAX 2365 4435

Email: informes.fondecyt@conicyt.cl

INFORME FINAL

PROYECTO FONDECYT REGULAR

OBJETIVOS

Cumplimiento de los Objetivos planteados en la etapa final, o pendientes de cumplir. Recuerde que en esta sección debe referirse a objetivos desarrollados, NO listar actividades desarrolladas.

N°	OBJETIVOS	CUMPLIMIENTO	FUNDAMENTO
1	Develop novel and particularized data mining techniques.	TOTAL	Como se explica en la sección de resultados, se desarrollaron técnicas novedosas de minería de datos aplicadas al contexto de agregación y sindicación de contenido y RSS. Se desarrolló un método de análisis de links (extracción de asociaciones) cuyos resultados fueron presentados en dos conferencia importantes con actas en ISI Proceedings. Se desarrolló un modelo de actualización de metadatos Web que se presentó en una publicación aceptada en una revista ISI de alto impacto en el área. También se desarrollaron métodos de análisis de uso de agregadores y contenido (texto) RSS. Uno de estos métodos, sobre mantención de directorio de recursos Web, se documentó en un paper enviado a revisión a un importante revista ISI del área. Otros métodos se encuentran documentados en borradores (adjuntos en sección de productos) y están preparando artículos para enviar a conferencias y revistas a la brevedad.

2	Develop proper representations of RSS data for data mining.	TOTAL	Se desarrolló un modelo de datos multidimensional de un agregador Web. El modelo registra información de uso (logs), RSS, canales sindicados, consultas, y otras entidades. El modelo se cargó con dos años de RSS extraído desde un "crawler" desarrollado previamente al proyecto, lo que involucra RSS sindicado por más de 20.000 canales y publicados diariamente en un agregador de RSS real (orbitando.com). Esto permitió obtener una base de datos multidimensional para hacer análisis e implementar métodos de minería de datos. Esta base de datos se utilizó para validar métodos desarrollados y realizar un completo estudio de la blogósfera Chilena que denominamos "Demografía de la Blogósfera Chilena". Adicionalmente se utilizó para extraer una red social de políticos mencionados en artículos RSS para analizar su presencia y centralidad.
3	Study data mining techniques in the context of data streams, the combination of text and semistructured data, and scarcity of computational resources.	TOTAL	El método de extracción de términos desarrollado (ver sección de resultados) ha significado procesar "streams" de datos y combinar análisis de texto e información semiestructurada presente en RSS. Sin embargo, todavía falta estudiar la eficiencia bajo escasos recursos computacionales (cuando las aplicaciones deben correrse en un software cliente). Adicionalmente, se ha estudiado el problema de segmentar ("clustering") RSS para detectar tópicos. Se ha desarrollado una versión del algoritmo Single Pass, que resuelve el problema para "streams". Se espera documentar y publicar este trabajo durante el segundo año del proyecto.
4	Improve applications such as ranking, filtering, information extraction in content syndication systems.	TOTAL	Como se explica en la sección de recursos, el problema de "rankear" y filtrar información ha sido central en el proyecto y se han generado resultados significativos al respecto. Se desarrolló un método para analizar sesiones Web con el objeto de mantener actualizado directorios de recursos Web. Este método involucra procesamiento de texto y datos que conforma un directorio Web y se enfoca en mejorar los rankings de recursos Web asociados a las categorías del directorio. El método para extraer asociaciones y caminos desde redes de recursos usando calce aproximado de patrones (explicado en la sección resultados) plantea contribuciones novedosas para los problemas de rankear y filtrar información organizada como redes (grafos).

Otro(s) aspecto(s) que Ud. considere importante(s) en la evaluación del cumplimiento de objetivos planteados en la propuesta original o en las modificaciones autorizadas por los Consejos.

Es importante destacar que al final del primer año el proyecto cambió de institución patrocinante y el investigador, al renunciar a su nombramiento en la institución anterior, debió abandonar la guía de al menos dos tesis relacionadas al proyecto, por lo cual no existieron tesis en el primer año del proyecto. En el segundo año del proyecto se hizo un esfuerzo y se logró desarrollar tres tesis de alumnos de la nueva institución patrocinante (dos de ellas a finalizar en Abril de 2010 y una en Mayo de 2010).

RESULTADOS OBTENIDOS:

Para cada uno de los objetivos específicos, describa o resuma los resultados. Relacione las publicaciones y /o manuscritos enviados a publicación con los objetivos específicos. En la sección Anexos incluya información adicional que considere pertinente para efectos de la evaluación.

La extensión máxima de esta sección es de 5 páginas (letra tamaño 10, Arial o Verdana).

Objetivo 1: Develop novel and particularized data mining techniques.

Los métodos de Minería de Datos en la Web (Minería Web) se clasifican en tres grupos: minería de estructura (analizar la estructura de redes de recursos), minería de uso (analizar el comportamiento de usuarios de la Web) y minería de contenido (analizar el contenido en texto de recursos Web). Seguiremos esta clasificación muy difundida en el área para presentar los resultados relacionados a este objetivo.

Minería de Estructura

Como se explica en la propuesta del proyecto, RSS es una familia de formatos para syndicar contenido Web (publicar contenido para que éste sea procesado automáticamente por terceros en la Web). RSS permite describir contenido como redes de recursos conectados por arcos marcados con propiedades (redes de información sindicada), lo que entrega una flexibilidad para representar desde artículos en blogs y medios tradicionales, hasta multimedia e información almacenada en bases de datos de grafos y relacionales. En este contexto, los sistemas que agregan contenido sindicado se pueden ver como integradores de información con estructura de grafo, publicada en múltiples fuentes de información, con esquemas (estructuras) diversos y autónomos. Un problema que surge en este contexto, es analizar o "minar" redes de información sindicada cuya estructura es desconocida por los sistemas y personas que las analizan.

En esta línea, se estudió el problema de extraer pares de recursos, llamados asociaciones, conectados por caminos que satisfacen determinados patrones en una red de información. Como es común en lenguajes de consulta sobre grafos, se estudio el caso donde los patrones se describen usando expresiones regulares. Cuando se desconoce la estructura de la información, los patrones pueden no tener sentido y sólo presentan una representación "aproximada" de las asociaciones que se quieren extraer. Para resolver este problema, se desarrolló un método de extracción y ranking de asociaciones en base a una semántica de consultas que permite calces aproximados entre patrones y caminos en la red de información. En el método, el patrón original se modifica incrementalmente, usando operaciones de edición. Cada nueva versión del patrón modificado permite extraer nuevas asociaciones aproximadas. Los resultados de este trabajo se documentaron en el artículo "**Ranking Approximate Answers to Semantic Web Queries**" (ver sección de productos), que fue presentado en la Conferencia Europea de la Web Semántica, versión 2009. En este artículo se explica el método con ejemplos prácticos (red de información de transporte entre ciudades) y se muestra un estudio algorítmico y de complejidad. Cabe destacar que el artículo fue aceptado entre 45 artículos de un total de 244 artículos recibidos por el comité de programa de la conferencia. La idea anterior se generalizó para considerar la extracción aproximada de caminos (es decir no sólo asociaciones sino los caminos que las conectan) que satisfacen patrones (expresiones regulares) desde redes de información. Los resultados de este trabajo se publicaron en el paper "**Finding Top-K Approximate Answers to Path Queries**", que fue presentado en la conferencia Flexible Query Answering Systems, versión 2009. En la actualidad se trabaja en una versión de revista del material publicado en las dos conferencias anteriores. El documento de trabajo de esta versión se titula "**Flexible Querying of Semantic Web Path Queries**" (ver sección resultados) y se espera enviar a Journal of Web Semantics en los próximos meses.

Por último, se estudio el problema de actualizar bases de metadatos para syndicar contenido en la Web: RDFS. El problema se estudió desde la perspectiva de actualizar una base de conocimiento. Se propusieron operadores lógicos para eliminar y agregar conocimiento. Se estudió la complejidad de implementar los operadores. Este trabajo se documento en el artículo titulado "**Updates in RDFS**" (ver sección de resultados y carta de aceptación en anexo). Una primera versión de este artículo fue enviado a la revista Journal of Web Semantics (Elsevier, ISI, Índice de impacto 3.023), siendo aceptada su publicación sujeto a incorporar comentarios de los revisores. Se está trabajando en terminar una nueva versión mejorada.

Minería de Uso

Otra línea de trabajo se centró en la extracción de patrones del comportamiento de usuarios en agregadores de RSS. En particular, se desarrolló un método para detectar contextos temporales (combinaciones de días de la semana y horas del día) donde surgen determinadas necesidades de información (representadas como temas). Para esto se tomó como punto de partida una idea previa del investigador responsable publicada como resumen (6 páginas) en SPIRE el año 2006. Se desarrolló en detalle el método expuesto en este artículo y se extendió para modelar contextos como regiones rectangulares en espacios categóricos de múltiples facetas. Este modelo se utilizó para extraer asociaciones entre contextos y temas y se realizaron experimentos con el registro de uso (log) de un buscador real. Este trabajo se documentó en el artículo "**Extracting interesting Web user Contexts**" (ver sección de productos). Se espera depurar y enviar este artículo a revisión a una revista de prestigio del área de Web mining.

Otro resultado relevante se documentó en el artículo "**Automatic Maintenance of Web Directories by Mining Web Browsing Data**" (ver sección productos) enviado a revisión a la revista Information Processing & Management (Elsevier, ISI, Índice de impacto 1.852). En este trabajo se propone un método para actualizar en forma continua un directorio de recursos Web, en base al análisis del comportamiento de los usuarios del directorio y sesiones de usuarios que visitan recursos posiblemente fuera del directorio. Como se ilustra en el artículo, el método tiene especial aplicación en la actualización de directorios de canales RSS. El artículo presenta el planteamiento teórico del método y experimentos basados en un directorio real de un motor de búsqueda (todo.cl), además de la implementación en un directorio y motor de búsqueda

Minería de Contenido

Una tercera línea de investigación se centró en el desarrollo de métodos de extracción de información en el contexto de RSS. En particular, se estudió el problema de extraer en forma automática nubes de términos (nubes de tags) desde colecciones de artículos RSS. Los tags son términos relevantes de distintos tipos que mezclan dimensiones distintas de la información. Las nubes de tags son interfaces muy populares para filtrar y clasificar artículos en agregadores RSS. Un problema esencial de las nubes de tags es su falta de estructura. Como ejemplo, tags que identifican personas, ciudades o países representan distintas dimensiones y éstas no se distinguen en las interfaces basadas en nubes de tags. Para solucionar este problema se aplicaron técnicas de áreas como Extracción de Información, Clasificación de Textos y Minería de Textos para extraer tags desde colección de artículos RSS y organizarlos en una estructura (compuesta por tipos de tags como personas, lugares, organizaciones, etc). En el artículo titulado "**Extracción de Nubes de Términos Categorizados desde RSS**" se documenta una primera versión del método. Se pretende depurar este trabajo integrando nuevos experimentos para completar un artículo para enviar a una conferencia relevante del área. Adicionalmente, se ha trabajado en el problema de segmentar RSS para detectar tópicos emergentes en un flujo de artículos que llegan a un agregador de contenido. Se ha desarrollado una versión del algoritmo "Single Pass" que se espera documentar en un artículo durante el segundo año del proyecto.

Objetivo 2: Develop proper representations for RSS data for Data Mining.

Se desarrolló un modelo multidimensional de datos que almacena información de contenido, uso y estructura (enlaces) acopiada por un agregador RSS. Este modelo se implementó en un data warehouse y se cargó con datos recolectados desde orbitando.com. La base de datos contiene, hasta el año 2008, un total de aproximadamente 19000 canales y 4650000 artículos. Los canales son en su mayoría chilenos (11846), seguidos por españoles (2508), argentinos (3082) y mexicanos (1660). Por su parte, los artículos provienen de blogs (2522592), noticias (920179), y contenidos de portales de contenidos: flickr, youtube, fotolog (1207380).

Este es un resultado novedoso sobre modelación de datos multidimensional y data warehouse aplicado al análisis de información asociada a un agregador RSS. Esta base de datos se utilizó para elaborar un estudio estadístico de la blogósfera y medios de prensa en la Web Chilena.

Este estudio está documentado en la tesis de magister del alumno Alvaro Chavez titulada "**Demografía de la Blogósfera Chilena**" (ver sección productos) ya finalizada como documento y en espera del examen final para su aprobación. Entre otros resultados, se presenta una estimación del tamaño de la blogósfera chilena para el año 2008 usando métodos estadísticos. Adicionalmente, se presenta una taxonomía de blogs, nubes de tópicos y análisis de comportamiento (publicaciones). En el estudio se contrastan estos aspectos entre regiones y región metropolitana y entre medios tradicionales y blogs. Durante Febrero de 2010 y en la actualidad se trabaja en actualizar este estudio para el año 2009 y publicarlo en algún medio de prensa con impacto a nivel nacional. Este trabajo se realizó con la ayuda del ingeniero Daniel Gómez y del profesor de la facultad de Gobierno de la UAI, Jorge Fábrega.

El data warehouse de RSS también se está aplicando en la tesis del alumno Felipe Mallea titulada "**Modelo de Extracción y Análisis de Redes Sociales**" (ver sección productos), donde se propone extraer la red asociaciones (coocurrencias) en artículos RSS de un universo de 300 políticos (ministros, senadores, diputados) con el objeto de estudiar cómo se comportan indicadores de centralidad en esa red social versus indicadores de presencia (número de menciones). En esta tesis también se propone y aplica un modelo de consultas analíticas para acceder a la red social en distintos niveles de abstracción (por ejemplo, a nivel de partido político, tipo de medio, etc.).

Objetivo 3: Study Data Mining Techniques in the context of data streams, the combination of text and semistructured data, and scarcity of computational resources.

Los métodos desarrollados y explicados en el objetivo 1 involucran el estudio de técnicas de minería de datos conocidas para el contexto de sindicación de contenido. Los resultados explicados en el Objetivo 1 también están relacionados a este objetivo. Por ejemplo, en el artículo "**Extracting interesting Web user Contexts**" se presenta una nueva clase de reglas de asociación (llamadas context-topic rules) que se aplican para extraer contextos interesantes desde logs Webs de agregadores y buscadores. En los artículos "**Ranking Approximate Answers to Semantic Web Queries**" y titula "**Flexible Querying of Semantic Web Path Queries**" se presenta una nueva técnica para extraer asociaciones relacionada al área de "link analysis". En el artículo "**Automatic Maintenance of Web Directories by Mining Web Browsing Data**" se presenta una nueva técnica para actualizar directories de recursos Web que se basa en el comportamiento de usuarios registrado en sesiones Web (a diferencia de la mayoría de las técnicas anteriores para este fin que se basan en el análisis del contenido de los recursos). Todas estas técnicas están extensamente explicadas en los distintos artículos (ver sección de productos). Por último la tesis de magister del alumno Claudio Días titulada "**Metodología de Crawling y Extracción de Datos para Comercio Electrónico en Sudamérica**" (ver sección productos) estudia el problema de monitorear eficientemente cambios de precios en productos publicados en la Web. Para esto se propone extraer desde series de tiempo de precio un modelo predictivo que prediga la frecuencia de actualización de información de precios para programar en forma más eficiente la recolección y extracción de esta información.

Objetivo 4: Improve applications such as ranking filtering, information extraction in content syndication systems.

Los métodos desarrollados y explicados en el objetivo 1 involucran mejoras de algoritmos de ranking (artículos "**Ranking Approximate Answers to Semantic Web Queries**" y "**Finding Top-K Approximate Answers to Path Queries**") y filtrado y extracción de información (artículo "**Extracción de Nubes de Términos Categorizados desde RSS**"). El método presentado en el artículo "**Automatic Maintenance of Web Directories by Mining Web Browsing Data**" se basa en la definición de un filtro y ranking de recursos asociados a las categorías de un directorio.

DESTAQUE OTROS LOGROS DEL PROYECTO TALES COMO:

- Estadías de investigación.
- Formación de recursos humanos exceptuando tesis ya informados.
- Actividades de difusión y/o extensión en la temática del proyecto.
- Cualquier otro logro no contemplado en los ítem anteriores y que Ud. quiera destacar.

La extensión máxima de esta sección es de 1 página (letra tamaño 10, Arial o Verdana).

Es importante destacar la colaboración internacional que se ha dado en relación al proyecto entre el investigador principal e investigado del Birkbeck College, Universidad de Londres y del London Knowledge Lab (Mark Levene, Peter Wood y Alexandra Poulouvasilis). El investigador principal realizó una visita académica de 10 días durante Septiembre de 2008 a Londres. Los profesores Peter Wood y Alexandra Poulouvasilis realizaron pasantías de investigación de una semana cada uno Santiago durante 2008. Estos profesores dictaron seminarios en el Departamento de Ciencias de la Computación de la Universidad de Chile y trabajaron con el investigador responsable. Estas pasantías fueron financiadas por el proyecto de la Royal Society titulado "**Flexible Querying of Network-Structured Data**", del cual el investigador responsable es codirector (ver anexo 3).

Por otra parte el investigador responsable ha mantenido una fructífera colaboración en distintos temas del proyecto con ex-alumnos del Departamento de Ciencias de la Computación de la Universidad de Chile, entre los cuales se encuentran Daniel Gómez, Cristián Serpell y Ricardo Zilleruelo.

En Mayo de 2008, el investigador dictó un seminario a profesores y alumnos de la facultad de Ingeniería de la Universidad Adolfo Ibáñez titulado "**Minería de Datos en Agregadores Web**". El material de esta presentación se adjunta en la sección de productos.

En Noviembre de 2008, el investigador responsable dictó un seminario titulado "**Inteligencia Artificial y la Web**" en el marco del Segundo taller de Tendencias Actuales en Ciencia e Ingeniería. Este es un curso de extensión ofrecido por la facultad de Ingeniería y Ciencias de la Universidad Adolfo Ibáñez y con amplia difusión en medios de prensa (ver anexo 2). El taller tuvo una numerosa concurrencia y se abordó en detalle el estado del arte de la sindicación de contenido en la Web. El material de la presentación se adjunta en la sección de productos.

Por último, es importante destacar la participación del investigador responsable en el **orbitando.com**, un agregadores de contenido de medios de prensa y blogs relacionados a Chile. El investigador participó activamente en el desarrollo de **tendenciaspoliticas.cl**, un portal Web pionero en el análisis y monitoreo de RSS sobre actores políticos. En particular, el investigador responsable ha participado en el desarrollo de técnicas de minería de datos para medir el impacto de actores políticos en blogs y medios digitales (ver anexo 5). Esto ha resultado en un fructífero intercambio de ideas que han enriquecido este proyecto Fondecyt.

Por último, este proyecto ha sido la base para la formación de un laboratorio denominado Social Media Institute, proyecto conjunto de las facultades de Gobierno e Ingeniería de la Universidad Adolfo Ibáñez. El data warehouse de RSS, obtenido de los canales de orbitando.com, será la base de datos fundamental de este laboratorio para desarrollar distintos estudios de los medios sociales y redes sociales en Chile.

Resumen:

Por sindicación de contenido nos referimos a la publicación continua de información en la Web por parte de sitios denominados canales cuyo objeto es que terceros la procesen y vuelvan a publicar o combinar (agregar) en forma automática. Este mecanismo que permite una mayor difusión de información, en particular asociados a medios, se ha difundido ampliamente en la Web gracias a una familia de formatos para publicar metadatos sobre el contenido sindicado, conocidos como RSS. En este proyecto se estudio la aplicación de técnicas de minería de texto y datos para mejorar sistemas que sindicán y consumen o agregan contenido sindicado. En particular, se desarrollaron métodos y algoritmos para organizar directorios de canales RSS, para analizar flujos RSS sindicado y extraer y analizar redes sociales y conexiones entre personas y entidades mencionadas en flujos de RSS. Se sentaron bases para estructurar grandes volúmenes de RSS y analizar esta información para mejorar distintas aplicaciones en el contexto de la sindicación de contenido.

PRODUCTOS

ARTÍCULOS

Para trabajos en Prensa/ Aceptados/Enviados adjunte copia de carta de aceptación o de recepción.

Nº : 1
Autor (a)(es/as) : Hurtado, CA; Mendoza M.
Nombre Completo de la Revista : Journal of Web Engineering
Título (Idioma original) : Automatic Maintenance of Web Directories by Mining Web Browsing Data
Indexación : ISI
ISSN :
Año : 2011
Vol. : 10
Nº : 2
Páginas : 20
Estado de la publicación a la fecha : Publicada
Otras Fuentes de financiamiento, si las hay :

Envía documento en papel : si

Archivo(s) Asociado(s) al artículo :

Journal_JWE.pdf

http://sial.fondecyt.cl/index.php/investigador/f4_articulos/descarga/6376500/1080672/2009/7857/1/

Nº : 2
Autor (a)(es/as) : Gutierrez
Nombre Completo de la Revista : Journal of Web Semantics
Título (Idioma original) : Updating RDFS Databases
Indexación : ISI
ISSN :
Año :
Vol. :
Nº :
Páginas :
Estado de la publicación a la fecha : Enviada
Otras Fuentes de financiamiento, si las hay :

Envía documento en papel : no

Archivo(s) Asociado(s) al artículo :

jswEnviadoPrevio.pdf

http://sial.fondecyt.cl/index.php/investigador/f4_articulos/descarga/6376500/1080672/2009/8030/1/

OTRAS PUBLICACIONES / PRODUCTOS

N° : 1
Autor (a)(es/as) : Carlos Hurtado, Mark Levene
Título (Idioma original) : Discovering Interesting Web User Contexts
Tipo de publicación o producto : Informe Técnico
ISBN :
Editor (es) (Libro o Capitulo de libros) :
Nombre de la editorial /Organización :
País : CHILE
Ciudad :
Fecha : Marzo - 2009
Año :
Vol. :
N° :
Páginas :
Otras Fuentes de financiamiento, si las hay :

Envía documento en papel : no
Archivo(s) Asociado(s) al artículo :
contextTopicRules.pdf
http://sial.fondecyt.cl/index.php/investigador/f4_otras_publicaciones/descarga/6376500/1080672/2009/2171/1/

N° : 2
Autor (a)(es/as) : C. Hurtado, A. Poulouvasilis, P. Wood
Título (Idioma original) : Flexible Querying of Semantic Web Path Queries
Tipo de publicación o producto : Monografía
ISBN :
Editor (es) (Libro o Capitulo de libros) :
Nombre de la editorial /Organización :
País : CHILE
Ciudad :
Fecha : Enero - 2010
Año :
Vol. :
N° :
Páginas :
Otras Fuentes de financiamiento, si las hay :

Envía documento en papel : no
Archivo(s) Asociado(s) al artículo :
jws_ongoing_(1).pdf
http://sial.fondecyt.cl/index.php/investigador/f4_otras_publicaciones/descarga/6376500/1080672/2009/2189/1/

Nº : 3
Autor (a)(es/as) : Carlos Hurtado, Cristián Serpell
Título (Idioma original) : Extracción de Nubes de Términos Categorizados
Tipo de publicación o producto : Informe Técnico
ISBN :
Editor (es) (Libro o Capítulo de libros) :
Nombre de la editorial /Organización :
País : CHILE
Ciudad :
Fecha : Enero - 2009
Año :
Vol. :
Nº :
Páginas :
Otras Fuentes de financiamiento, si las hay :

Envía documento en papel : no

Archivo(s) Asociado(s) al artículo :

extraccionTerminos23.pdf

http://sial.fondecyt.cl/index.php/investigador/f4_otras_publicaciones/descarga/6376500/1080672/2009/2313/1/

Nº : 4
Autor (a)(es/as) : Carlos Hurtado
Título (Idioma original) : Minería de Datos en Agregadores Web
Tipo de publicación o producto : Seminario/Taller/Curso
ISBN :
Editor (es) (Libro o Capítulo de libros) :
Nombre de la editorial /Organización :
País : CHILE
Ciudad :
Fecha : Mayo - 2008
Año :
Vol. :
Nº :
Páginas :
Otras Fuentes de financiamiento, si las hay :

Envía documento en papel : no

Archivo(s) Asociado(s) al artículo :

CharlaMineriaDeDatosEnAgregadoresWeb.pdf

http://sial.fondecyt.cl/index.php/investigador/f4_otras_publicaciones/descarga/6376500/1080672/2009/2314/1/

Nº : 5
Autor (a)(es/as) : Carlos Hurtado
Título (Idioma original) : Inteligencia Artificial y la Web
Tipo de publicación o producto : Seminario/Taller/Curso
ISBN :
Editor (es) (Libro o Capítulo de libros) :
Nombre de la editorial /Organización :
País : CHILE
Ciudad : Santiago
Fecha : Noviembre - 2008
Año :
Vol. :
Nº :
Páginas :
Otras Fuentes de financiamiento, si las hay :

Envía documento en papel : no

Archivo(s) Asociado(s) al artículo :

charlaTendenciasCienciaYtecnologia.pdf

http://sial.fondecyt.cl/index.php/investigador/f4_otras_publicaciones/descarga/6376500/1080672/2009/2316/1/

CONGRESOS

Nº : 1
Autor (a)(es/as) : Hurtado, CA.; Poulouvasilis A.; Wood PT.
Título (Idioma original) : Finding Top-K Approximate Answers to Path Queries
Nombre del Congreso : Flexible Query Answering Systems, 8th International Conference, FQAS 2009
País : DINAMARCA
Ciudad : Roskilde
Fecha Inicio : 26/10/2009
Fecha Término : 28/10/2009
Nombre Publicación : Lecture Notes in Computer Science
Año : 2009
Vol. : 5822
Nº :
Páginas : 465-476
Envía documento en papel : no
Archivo Asociado :
fqas_2009.pdf
http://sial.fondecyt.cl/index.php/investigador/f4_congresos/descarga/6376500/1080672/2009/11996/1/

N° : 2
Autor (a)(es/as) : Hurtado, CA.; Poulouvassilis A.; Wood PT.
Título (Idioma original) : Ranking Approximate Answers to Semantic Web Queries
Nombre del Congreso : 6th European Semantic Web Conference, ESWC 2009
País : GRECIA
Ciudad : Heraklion
Fecha Inicio : 31/05/2009
Fecha Término : 04/06/2009
Nombre Publicación : Lecture Notes in Computer Science
Año : 2009
Vol. : 5554
N° :
Páginas : 263-277
Envía documento en papel : no
Archivo Asociado :
flex-reg-exp.pdf
http://sial.fondecyt.cl/index.php/investigador/f4_congresos/descarga/6376500/1080672/2009/11998/1/

TESIS/MEMORIAS

N° : 1
Título de Tesis : Modelo de extracción y Análisis de Redes Sociales
Nombre y Apellidos del(de la) Alumno(a) : Felipe Mallea
Nombre y Apellidos del(de la) Tutor(a) : Carlos Hurtado
Título Grado : Magister
Institución : Facultad de Ingeniería Universidad Adolfo Ibáñez
País : CHILE
Ciudad : Santiago
Estado de Tesis : En Ejecución
Fecha Inicio : 01/03/2009
Fecha Término : 30/04/2010
Envía documento en papel : no
Archivo Asociado :
TesisFelipeMallea.pdf
http://sial.fondecyt.cl/index.php/investigador/f4_tesis_memorias/descarga/6376500/1080672/2009/5940/1/

N° : 2
Título de Tesis : Metodología de Crawling y Extracción de Datos para Comercio Electrónico en Sudamérica
Nombre y Apellidos del(de la) Alumno(a) : Claudio Díaz
Nombre y Apellidos del(de la) Tutor(a) : Carlos Hurtado
Título Grado : Magister
Institución : Facultad de Ingeniería Universidad Adolfo Ibáñez

País : CHILE
Ciudad : Santiago
Estado de Tesis : En Ejecución
Fecha Inicio : 01/06/2009
Fecha Término : 30/06/2010
Envía documento en papel : no
Archivo Asociado :
TesisClaudioDiaz.pdf
http://sial.fondecyt.cl/index.php/investigador/f4_tesis_memorias/descarga/6376500/1080672/2009/5941/1/

Nº : 3
Título de Tesis : Demografía de la Blogósfera Chilena
Nombre y Apellidos del(de la) Alumno(a) : Alvaro Chávez Campusano
Nombre y Apellidos del(de la) Tutor(a) : Carlos Hurtado
Título Grado : Magister
Institución :
País : CHILE
Ciudad : Santiago
Estado de Tesis : En Ejecución
Fecha Inicio : 01/03/2009
Fecha Término : 30/04/2010
Envía documento en papel : no
Archivo Asociado :
TesisALVAROCHAVEZVERSIONEmpastada.pdf
http://sial.fondecyt.cl/index.php/investigador/f4_tesis_memorias/descarga/6376500/1080672/2009/6213/1/

ANEXOS

A continuación se detallan los anexos físicos/papel que no se incluyen en el informe en formato PDF.

--