



**Universidad CENFOTEC
Maestría en Bases de Datos
Proyecto de Investigación**

**Aumento en la Productividad de los
Desarrolladores por Medio de un Sistema de
Administración de Conocimiento**

Autor: Diego Alfaro Bergueiro

Tutor: Eduardo Castro

27/04/2013

Contenido

Resumen.....	5
Introducción.....	7
Objetivo General	9
Objetivos Específicos.....	9
Definición del Problema.....	10
Alcances.....	11
Limitaciones	12
Metodología.....	13
Cronograma.....	15
Justificación	16
Antecedentes y Proyectos Similares	19
Justificación de Decisiones de Diseño.....	20
Algoritmo de Minería de Datos.....	27
Ejemplo de consulta	28
Análisis y Discusión de Resultados.....	30
Plan de Inversión del Proyecto.....	32
Conclusiones y Trabajos Futuros.....	33
Anexos.....	35
Diagrama Entidad – Relación	36
Páginas del Sistema.....	37
Diccionario de Datos	44
Tablas	44
Vistas	49
Procedimientos Almacenados.....	50
Funciones	61
Bibliografía	63

Resumen

Ante el reto de asegurar la distribución de conocimiento entre sus empleados al cual se enfrentan la mayoría, sino todas, las empresas transnacionales y otras empresas que por su tamaño les sea necesario distribuir su negocio en sedes geográficamente dispersas, surge la necesidad de buscar soluciones eficientes y sencillas, que apelen a los empleados y les provoquen utilizarlas. En el caso de los desarrolladores de software, se puede decir que se da un “caso especial” porque se trata de profesionales que generalmente están presionados a tener listos sus proyectos siguiendo una estricta agenda. A pesar de haber en internet gran cantidad de foros disponibles, donde los programadores pueden exponer sus dudas y/o buscar respuestas que se les haya brindado a otras personas con su misma duda, puede haber diversas razones que motiven a las empresas restringir el uso de estos foros públicos.

Como alternativa a algunas de esas razones, ya sean seguridad de información o un deseo de incrementar la colaboración entre los desarrolladores o programadores en las diferentes regiones geográficas donde la compañía Intel tiene sedes, se presenta en este proyecto un sistema de administración de conocimiento basado en algunas técnicas de inteligencia de negocios y aprovechando las facilidades brindadas por SQL Server 2012. La solución utiliza no solo herramientas “tradicionales” de ese producto como lo es el SQL Server Analysis Services (SSAS), si no también herramientas nuevas como las consultas semánticas. En síntesis, se creó una herramienta que permite a los desarrolladores plantear sus dudas o problemas y recibir soluciones (o proponerlas). El producto final utiliza un modelo de minería de datos implementado en SSAS para categorizar los datos insertados por los usuarios, y se combina esto con la búsqueda semántica para reducir el conjunto de datos a la hora de hacer búsquedas y a la vez realizar consultas avanzadas sobre los datos.

Introducción

La colaboración entre los empleados de la compañía es un tema que ha cobrado mucha fuerza en los últimos meses dentro de Intel. Históricamente, se han implementado varias soluciones al dilema de "cómo compartir información entre empleados de diferentes regiones geográficas". Por lo general, se implementan soluciones locales en cada departamento para transferir conocimientos. Por ejemplo: los miembros de un mismo grupo de desarrolladores se pueden poner de acuerdo para hacer presentaciones a sus compañeros sobre temas específicos. Similarmente, algunos grupos llevan a cabo reuniones semanales o mensuales donde se discuten temas importantes para todo el equipo.

A un nivel más macro, se ha implementado el uso de diferentes sitios en la intranet para facilitar la comunicación y colaboración entre empleados. Entre los más notables, se cuenta con un sitio donde se pueden crear foros con temas específicos. Los empleados de cualquier fábrica u oficina de Intel se pueden afiliar a dichos foros e iniciar conversaciones. Otro sitio notable es de estilo wiki, donde los empleados pueden crear secciones para temas específicos y postear información al respecto. Ambos sistemas son valiosos desde el punto de vista de compartir información, pero no han gozado de un buen nivel de aceptación por parte de los desarrolladores en términos de utilización, específicamente para compartir conocimiento sobre errores recibidos y las soluciones empleadas.

Adicionalmente, dentro de la compañía se realizan varias conferencias técnicas de manera anual. El enfoque de dichas conferencias es muy variado: desde temas muy específicos al área de manufactura hasta conferencias para desarrolladores. Estas conferencias se utilizan para compartir conocimiento entre empleados de todo el mundo. Sin embargo, son eventos que se realizan solo una vez al año, al cual no todos los empleados tienen acceso para o interés en participar. Adicionalmente, los proyectos presentados tienen un ámbito muy delimitado. No son espacios donde un programador pueda buscar respuestas a errores específicos que puede encontrar al desarrollar, sino donde podría encontrar información sobre diferentes proyectos o tecnologías que se utilizan en otras áreas de Intel.

Ante dicho panorama, el autor de este proyecto se planteó la incógnita de si será posible mejorar dicho ambiente al proveer una herramienta que permita a los desarrolladores de la compañía mejorar su colaboración entre ellos.

Objetivo General

Planear y desarrollar una estrategia de administración del conocimiento, orientado a desarrolladores, que permita a los programadores de Intel aumentar su productividad a la hora de desarrollar sistemas al contar con un proceso de gestión de conocimiento integrado.

Objetivos Específicos

1. Identificar esfuerzos realizados, dentro de Intel, por equipos dedicados al desarrollo de sistemas para transmitir el conocimiento adquirido por sus miembros al solucionar errores en sus aplicaciones o sistemas.
2. Identificar métricas utilizadas por los equipos de desarrollo en Intel para medir la productividad de los desarrolladores, así como los mecanismos para tomar dichas medidas.
3. Desarrollar e implementar un sistema de gestión de conocimiento que permita a los desarrolladores ingresar investigaciones realizadas o errores recibidos, así como las soluciones aplicadas o hallazgos en cada situación para que estas puedan ser consultadas por otros desarrolladores.
4. Desarrollar una estrategia para la promoción del sistema de gestión de conocimiento desarrollado.

Definición del Problema

Durante el proceso de desarrollo de un sistema o aplicación, un desarrollador frecuentemente encuentra errores con mensajes crípticos que no indican claramente qué es exactamente el error o cuál podría ser la solución. Similarmente, durante su carrera, un desarrollador puede necesitar implementar una nueva funcionalidad con la cual no está familiarizado, o tiene que trabajar con un lenguaje de programación que no maneja en su totalidad.

Al encontrarse ante alguna de estas situaciones, el programador se ve forzado a preguntar a sus compañeros o buscar soluciones alternas. En el mejor de los casos, alguno de sus compañeros le podrá indicar la solución más práctica; de lo contrario, el desarrollador debe comenzar a investigar por sus propios medios ya sea consultando libros o navegando en internet, al no tener un repositorio centralizado o un sistema de administración de conocimiento.

Indiferentemente de cuál de los dos escenarios se dé, el conocimiento entre desarrolladores se pasa de manera rudimentaria e informal (si es que se pasa). A pesar de la existencia de wikis, blogs y foros dentro y fuera de la compañía, los desarrolladores no utilizan una herramienta centralizada para compartir el conocimiento adquirido al buscar soluciones a errores de programación, sino que buscan respuestas en diferentes lugares hasta encontrarlas, lo cual puede implicar una inversión de tiempo significativa.

Alcances

En las primeras etapas de este proyecto, se lograron identificar algunas métricas utilizadas para medir la eficiencia de los programadores, así como algunos procesos utilizados por diferentes equipos de desarrolladores para compartir conocimiento dentro de la empresa. Al finalizar la etapa de desarrollo de la solución, se cuenta con una herramienta web capaz de realizar las siguientes tareas:

- 1) Crear plantillas de aplicaciones con características como el motor de base de datos que utiliza, el sistema operativo del servidor donde reside y el ambiente y lenguaje de programación utilizados para desarrollar dicha herramienta, entre otras.
- 2) Ingresar temas investigados (conocidos como “Researched Items” dentro del sistema), así como soluciones (“findings”) que los usuarios pueden proponer.
- 3) Buscar información en tres modalidades diferentes:
 - *Ítem específico*: buscar un tema investigado utilizando su identificador.
 - *Básica*: búsqueda semántica basada en palabras claves ingresadas por el usuario.
 - *Avanzada*: permite hacer una búsqueda semántica pero permite refinarla al indicar características de la aplicación o el ambiente en el que se está desarrollando.
- 4) Tener acceso rápido y directo a los últimos temas que se han ingresado (ya sea por la misma persona o por el resto de los usuarios).
- 5) Competir con otros usuarios para determinar quién es el mayor colaborador del sitio.
- 6) Presentar el proyecto en una conferencia técnica dentro de la compañía.

Limitaciones

El sistema desarrollado cumple con las especificaciones generales que se plantearon al inicio del proyecto, por lo tanto, en el tema de funcionalidad no hay limitaciones con respecto al ámbito original de proyecto. Sin embargo, durante la implementación del motor de búsqueda avanzado sí se encontró una limitación significativa, específicamente relacionada con el modelo de minería de datos. Originalmente se planeó utilizar criterios como: las arquitecturas de software, hardware y bases de datos utilizadas, así como los componentes de terceros involucrados en el desarrollo de la aplicación entre otros. A diferencia de las otras características, como el motor de base de datos utilizado o el lenguaje de programación, estas no tienen una relación 1:1 (la herramienta pone la restricción de solo poder asociar un motor o un lenguaje a cada aplicación), estas últimas se manejan en relaciones n:n (un componente o una arquitectura puede ser utilizado por n aplicaciones, y una aplicación puede utilizar n componentes o arquitecturas).

A la hora de diseñar el modelo de minería de datos para la búsqueda avanzada, se intentó agregar estas características que funcionan en relaciones n:n como tablas anidadas dentro de la estructura de minería. Desafortunadamente, no se logró completar la implementación exitosa de este diseño, por lo cual se decidió omitir dichas características y utilizar sólo aquellas que tienen una relación 1:1. Esta decisión se tomó al considerar la fecha de entrega del proyecto y la cantidad de trabajo y tiempo que tomaría re-diseñar el modelo de minería de datos y posiblemente la base de datos.

Adicionalmente, y también por consideraciones de tiempo, a pesar de utilizar la metodología de LEAN/Six Sigma como base para el desarrollo del proyecto, se descartó la etapa de *Control* con la cual normalmente se cierran los proyectos. La razón de esta decisión es la cantidad de tiempo que tomará proponer la utilización de la herramienta y la aceptación de la misma por los usuarios. Debido a esto, resultará imposible establecer y tomar mediciones de control antes de la fecha de entrega.

Metodología

Durante el desarrollo de este proyecto, en sus diferentes etapas, se utilizarán diferentes metodologías. Entre ellas:

Entrevistas Personales

Las entrevistas personales se utilizaron para entender mejor el entorno actual. Por medio de estas, recogió información sobre lo que se está haciendo hoy en día para gestionar el conocimiento en los equipos o departamentos enfocados en programar, sino también entender cómo se mide la productividad de dichos desarrolladores. Adicionalmente, las entrevistas se utilizaron también como medio de contactar usuarios potenciales. En este caso la intención de la entrevista no solo era identificar procesos actuales, sino también proponer la metodología en estudio como una mejora y así conseguir una base de usuarios más completa.

Encuestas

Las encuestas se utilizaron como el principal método de recopilar información de los usuarios, como medir la cantidad de tiempo que cada uno de ellos dedica a buscar soluciones a errores en su código en la intra o internet.

LEAN / Six Sigma

Intel ha adoptado la combinación de las metodologías LEAN y Six Sigma como uno de los principales métodos para depurar sus procesos y eliminar el desperdicio. Siguiendo esa línea, dado que esta iniciativa busca reducir el tiempo desperdiciado por desarrolladores en buscar soluciones en internet, se adoptaron estas metodologías como la base del proyecto. Entre las ventajas que la metodología LEAN ofrece, está la estructura: se trata de un proceso de cinco etapas las cuales se van a seguir para desarrollar este proyecto. Esto se puede ver reflejado en el cronograma.

En la primera etapa, se *Definió* el problema que a atacar. Esta definición fue más allá de la definición del problema establecida en este documento, y se procedió a identificar métricas así como las herramientas para tomar las respectivas mediciones. Una vez definido el proyecto, se procedió a *Medir*. En esta etapa se tomaron las mediciones para calcular las métricas

identificadas en la definición. Posteriormente, se pasó a la etapa de Análisis, donde se analizaron los datos recopilados durante la etapa de medición. Basado en este análisis se determinó la mejor forma de solucionar el problema. Una vez identificadas las posibles soluciones, se pasó a la etapa de Implementación. En esta etapa se desarrolló e implementó la solución seleccionada en la etapa de análisis. Finalmente, se llega a la etapa de Control. En esta etapa se toman mediciones nuevamente, para comprobar que las soluciones implementadas han sido efectivas y que el proceso ha sido mejorado.

En el caso de este proyecto, la etapa de control no va a formar parte del cronograma, pues antes de llegar a una etapa donde se pueda demostrar el efecto que el sistema de administración de conocimiento puede tener sobre el tiempo dedicado a investigar, se tiene que pasar por un proceso de promoción y comunicación. Adicionalmente, los usuarios deben comenzar gradualmente a ingresar información, y puede pasar una cantidad significativa de tiempo antes de poder tener un conjunto de datos sustancial para realizar análisis y pruebas de control.

Conferencias Técnicas

Dentro de Intel, se realizan diferentes conferencias técnicas a lo largo del año. Uno de los propósitos de estas conferencias es facilitar la tarea de compartir información entre diferentes empleados de la compañía. Como se mencionó anteriormente, cada una tiene su tema principal, ya sea desarrollo de sistemas o mejora de procesos. En las primeras etapas de este proyecto, se utilizó dichas conferencias como una forma de promocionar la adopción de la herramienta y conseguir una cantidad significativa de usuarios iniciales. En etapas posteriores al desarrollo, se planea utilizar dichas conferencias nuevamente para promocionar la herramienta y así conseguir una base de usuarios más amplia.

Cronograma

Task Name	Duration	Start	Finish	Predecessors
Definir	20 days?	Mon 4/9/12	Fri 5/4/12	
Investigar mediciones de productividad usadas en Intel	3 days?	Mon 4/9/12	Wed 4/11/12	
Investigar métodos de gestión de conocimiento usados en Intel	5 days	Thu 4/12/12	Wed 4/18/12	
Determinar métodos para medir la utilización del tiempo de los desarrolladores	4 days?	Thu 4/19/12	Tue 4/24/12	
Buscar usuarios potenciales	20 days?	Mon 4/9/12	Fri 5/4/12	
Levantar requerimientos para el Sistema de Gestión de Conocimiento	10 days?	Mon 4/23/12	Fri 5/4/12	
Medir	30 days?	Mon 5/7/12	Fri 6/15/12	1
Implementar un método de captura de datos para medir la distribución del tiempo productivo de los desarrolladores	5 days?	Mon 5/7/12	Fri 5/11/12	
Coordinar el grupo de usuarios que van a colaborar con las mediciones	6 days?	Wed 5/9/12	Wed 5/16/12	
Medir la utilización del tiempo de desarrolladores en diferentes equipos y diferentes regiones geográficas	25 days?	Mon 5/14/12	Fri 6/15/12	8
Analizar	30 days?	Mon 6/4/12	Fri 7/13/12	
Identificar patrones de distribución del tiempo de los desarrolladores	8 days?	Mon 6/4/12	Wed 6/13/12	
Identificar y estudiar los recursos utilizados con mayor frecuencia por los desarrolladores	8 days?	Wed 6/13/12	Fri 6/22/12	
Revisar y afinar la lista de requerimientos para el Sistema de Gestión de Conocimiento	10 days?	Mon 6/25/12	Fri 7/6/12	12,13
Identificar y analizar las razones por las cuales los usuarios utilizaron con mayor frecuencia recursos específicos	15 days?	Mon 6/25/12	Fri 7/13/12	12,13
Implementar	165 days?	Mon 7/16/12	Fri 3/1/13	11
Diseñar el Sistema de Gestión de Conocimiento	78 days?	Mon 7/16/12	Wed 10/31/12	
Programar y probar el Sistema de Gestión de Conocimiento	87 days?	Thu 11/1/12	Fri 3/1/13	17
Planear la estrategia de promoción y comunicación de la iniciativa	35 days?	Mon 9/3/12	Fri 10/19/12	
Comunicación y Promoción	215 days	Mon 3/4/13	Fri 12/27/13	18,19
Comunicar a diferentes organizaciones enfocadas al desarrollo de sistemas acerca del sistema de gestión de conocimiento	310 days?	Mon 10/22/12	Fri 12/27/13	
Documentación del Proyecto	275 days?	Mon 4/9/12	Fri 4/26/13	

Justificación

Cuando se habla de proyectos de desarrollo de sistemas, se aplica la máxima de que “el tiempo es oro”. Un atraso en la programación de un sistema puede tener consecuencias bastante considerables. Por ejemplo: atrasos en el desarrollo e implementación de un sistema para un manejo más eficiente del inventario en una fábrica puede atrasar proyectos de mejora en el manejo de los activos de una compañía y provocar que la compañía siga perdiendo dinero por el manejo ineficiente de su inventario. Conforme han pasado los años y ha evolucionado la administración de proyectos relacionados con Software, se han implementado diferentes metodologías que han buscado organizar o estructurar el proceso de desarrollo, muchas veces con el propósito de hacer que el desarrollo de los sistemas suceda de una manera más fluida y ordenada. Entre dichas metodologías podemos resaltar el modelo en cascada, la generación de prototipos, el desarrollo incremental, el desarrollo en espiral y la metodología Ágil. Estas metodologías tienen muchas similitudes entre ellas, pero la más relevante para este proyecto es “la búsqueda de una mayor eficiencia” en el proceso de desarrollar un sistema.

A pesar de que cada metodología tienen sus fortalezas, hay factores en el desarrollo de Software que, por más completa que sea una metodología, no se pueden controlar. Por ejemplo: estas metodologías proveen una estructura que se puede seguir para agilizar el proyecto y sus diferentes etapas, pero dichas mejoras en términos de agilidad están sujetas a que los desarrolladores puedan cumplir con sus tareas sin atrasos. Si un desarrollador tiene una semana de tiempo para completar un módulo determinado, y en el desarrollo de este encuentra un error al compilar su código, no hay nada en estas metodologías que ayuden al desarrollador a solucionar su dilema en la semana que tiene de tiempo para completar el módulo. Similarmente, si un desarrollador está trabajando con un lenguaje de programación con el cual no está completamente familiarizado, o necesita trabajar con funcionalidades diferentes a las que está acostumbrado, el desarrollador se ve en una situación donde la metodología seleccionada para trabajar el proyecto le va a ser de muy poca utilidad.

En estos casos, el desarrollador tiene que buscar recursos que le ayuden a solucionar su problema o error, ya sea en línea (Google, foros técnicos, etc.), consultando libros (manuales de referencia, manuales de certificación, etc.) o consultando a sus compañeros. Por medio de entrevistas personales con varios gerentes de equipos enfocados al desarrollo de sistema, se identificó la falta de un sistema integrado, apropiado para uso interno de la compañía, que permitiera compartir conocimiento entre los desarrolladores de la empresa. A pesar de que

existen gran cantidad de foros de este tipo en internet, por motivos de privacidad de información y de protección de información sensible para la compañía, no se pueden utilizar dichos sitios libremente. Adicionalmente, se cuentan con varias herramientas para colaboración dentro de la compañía, en la forma de blogs, sitios de Sharepoint e incluso una red social. Lamentablemente, ninguna de estas herramientas presenta una alternativa fácil de usar con el propósito de facilitar conocimiento que muchas veces se necesita inmediatamente para no sufrir más atrasos.

En estas entrevistas también se averiguó que no hay un conjunto estándar de métricas dentro de Intel que permita medir la eficiencia de los desarrolladores. Algunos equipos utilizan como métricas la cantidad de errores ("bugs" o "pulgas") en sus sistemas, mientras que otros aprovechan que metodologías como la programación orientada a pruebas ("Test-Driven Development") les facilitan sacar métricas como por ejemplo la cantidad de pruebas exitosas. Dado esto, fue necesario hacer un poco de trabajo de campo para lograr entender cuánto tiempo dedican los desarrolladores a investigar cómo hacer o arreglar algo para entender el impacto que esto tiene sobre su desempeño.

Se concertó una entrevista personal a un desarrollador de sistemas de la compañía, y en esta se averiguó que en promedio él invierte entre 12 y 15 horas por mes navegando por internet, buscando soluciones a errores recibidos al momento de compilar diferentes partes de su código. Trabajando con este estimado, se mantendría un promedio de 12 horas por mes y se hablaría de 144 horas por desarrollador por año invertidas en búsqueda de soluciones o de información. Adicionalmente, si se escala esta estadística a la cantidad total de desarrolladores que tiene Intel, se llegaría a un número bastante significativo. Por ejemplo: si se considera un grupo de 100 desarrolladores, ya se hablaría de alrededor de 14 mil horas hombre invertidas en la búsqueda de soluciones.

Para corroborar este estimado, se desarrolló una herramienta que permitiera a desarrolladores de diferentes organizaciones dentro de la compañía llevar un registro de la cantidad de veces que han tenido que buscar información en línea, así como la cantidad de sitios visitados y los sitios específicos visitados, ya sea para solucionar errores recibidos o para buscar información sobre "cómo hacer algo". En un lapso de seis semanas, trece desarrolladores de diferentes regiones geográficas y de diferentes organizaciones contribuyeron con información sobre búsquedas o investigaciones realizadas en línea. Durante estas semanas, se mantuvo una media de 3.21 horas invertidas en total por los desarrolladores, y una media de 5 sitios visitados por

semana. Al realizar la proyección con estos valores, se valida el estimado inicial. La media de 3.21 horas por semana implicaría un aproximado de 12.84 horas por mes por desarrollador, y se convierte en un aproximado de 154.16 horas por desarrollador por año destinadas a la tarea de investigar en línea.

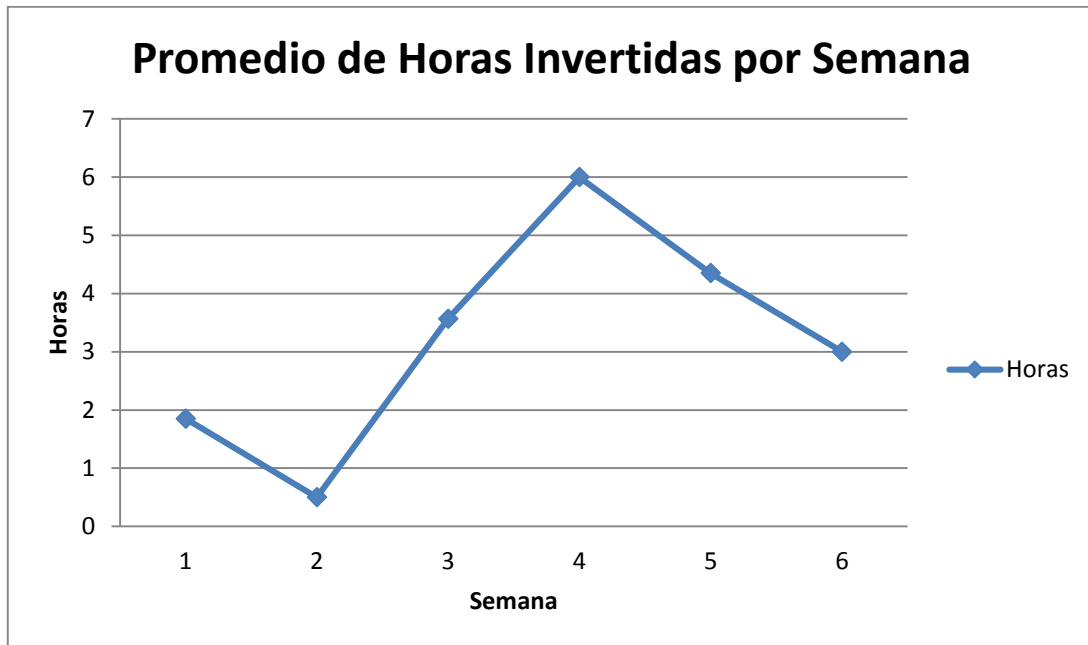


Gráfico 1: Horas invertidas en promedio por semana en Investigaciones en línea.

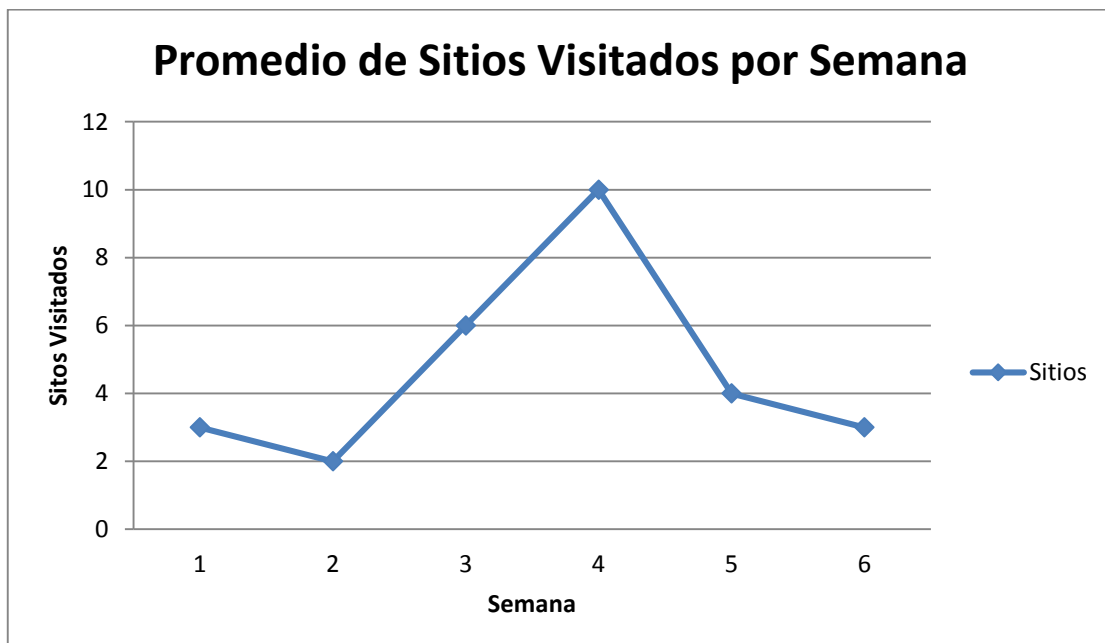


Gráfico 2: Cantidad de sitios visitados en promedio por semana

Desde el punto de vista de la metodología LEAN, el tiempo invertido en dicha consulta es un desperdicio, pues la tarea de programar se ve interrumpida por una actividad que no agrega valor al proceso de desarrollo. Aún sin entrar de lleno a hacer un mapa del proceso de desarrollo de sistemas, resulta evidente que una de las principales fuentes de desperdicio es que los desarrolladores tengan que navegar por diferentes sitios web y/o foros para encontrar la solución a los problemas que enfrente al programar, especialmente cuando algún otro desarrollador puede tener la respuesta o la solución.

A pesar de poder cuantificar el impacto que podría tener un sistema de gestión de conocimiento en horas que un desarrollador se ahorraría, es necesario también complementar la implementación de este sistema con una estrategia para gestionar el conocimiento adquirido. Dicha estrategia es necesaria para garantizar que el desarrollo del sistema no llegue a hacer "un sistema más" sino que forma parte de un marco de trabajo más completo, y sea adoptado por los desarrolladores de sistemas en toda la compañía en lugar de solo unos cuantos.

Antecedentes y Proyectos Similares

Herramientas Internas de Colaboración:

Como se mencionó anteriormente en el documento, a nivel corporativo hay varias herramientas de colaboración que pueden ser utilizadas para compartir información. Se cuenta con un sistema de chat interno y un sistema de correo electrónico internos, así como de varios sitios de Microsoft SharePoint que pueden ser utilizados para compartir documentos, entre otro tipo de información.

Herramientas Comerciales:

A nivel comercial, se ofrecen diferentes herramientas que buscan específicamente facilitar la administración de conocimiento, o herramientas en las cuales la administración de conocimiento es un módulo de la misma. Algunos ejemplos son XchangeMatrix¹ y ServiceNow². La primera herramienta muestra ciertas limitaciones en cuanto a aplicabilidad en el ámbito definido en este proyecto, pues se ofrece como una solución para el manejo de conocimiento de toda la empresa, incluyendo la relación con sus clientes. Con respecto a ServiceNow, la

¹ <http://www.betapersei.com/xchangematrix/>

² <http://www.servicenow.com/>

administración de conocimiento se incluye en ella al ser esta una herramienta pensada en la implementación de ITIL, y el manejo del conocimiento parte importante de dicha metodología.

Buscadores Web:

Se podría considerar a los buscadores web, como Google por ejemplo, como un gran sistema gestor de conocimiento. Esto porque resulta muy fácil encontrar artículos, libros y demás medios informativos con suma facilidad. Dejando de lado las implicaciones de seguridad de información y de colaboración, la iniciativa descrita en este proyecto se diferencia de dicho buscador porque en ella se utiliza la minería de datos (clustering, específicamente), para generar un índice de similitud que posteriormente se utiliza para filtrar resultados. Adicionalmente, este proyecto utiliza búsquedas semánticas, por lo que “entiende” o “interpreta” lo que el usuario quiere buscar. En el caso de Google, se utiliza un algoritmo para asignar un peso al origen de la información, el cual busca filtrar aquellos resultados que no vengan de fuentes confiables. Posteriormente, se analiza el texto de búsqueda y se realiza la consulta en base al texto ingresado, no al contexto o a la intención del usuario.

A nivel interno, en la compañía se utilizan herramientas de colaboración y una herramienta comercial aparte para la administración de conocimiento en general. Sin embargo, no hay nada orientado específicamente a los desarrolladores. Adicionalmente, el SQL Server 2012 es una tecnología que apenas se está comenzando a utilizar, por lo cual su uso no se ha propagado aún. Finalmente, la minería de datos se está comenzando a utilizar en algunas investigaciones a nivel local, pero no es una práctica extensa todavía.

Justificación de Decisiones de Diseño

Diseño de la Base de Datos:

Siguiendo las mejores prácticas de diseño de bases de datos, los primeros objetos que se crearon en ella fueron los esquemas. La base de datos desarrollada cuenta con tres diferentes esquemas, en donde se agrupan los objetos dependiendo de su función: producción, referencia o “staging”. El esquema de referencia contiene los metadatos principalmente, toda aquella información de catálogo utilizada para trabajar con las plantillas. Por lo tanto, en este esquema se encuentran las tablas que contienen las plantillas y las características de cada una. El esquema de producción se utiliza para almacenar los objetos relacionados con los hallazgos (“findings”, “research items”

y los respectivos comentarios). Finalmente, el esquema de “staging” contiene la tabla utilizada en el proceso de consulta automatizada al modelo de minería de datos. En esta tabla se almacena el resultado de la consulta para luego ser aplicada a la tabla de las plantillas en producción.

La mayoría de las relaciones en el diseño de la base de datos son del tipo 1:n, por ejemplo: un motor de base de datos específico puede ser utilizado por diferentes plantillas, pero a una plantilla sólo se le permite especificar un motor de bases de datos. La excepción a esto son las tablas que manejan las arquitecturas de hardware, bases de datos y software, así como las que manejan los componentes de terceros utilizados. En estos casos se utilizan tablas intermedias para habilitar las relaciones n:n entre las características y las plantillas, de manera que una plantilla pueda utilizar diferentes componentes de terceros y un componente específico puede ser utilizado por varias plantillas.

Para habilitar la búsqueda semántica, fue necesario crear un “Full-Text Catalog”, o catálogo de texto. Este componente de la base de datos se configuró para ser utilizado sobre los datos ingresados a la tabla “ResearchedItems”. En esta tabla se creó un índice de texto (Full-Text Index) que incluye las columnas “ResearchDescription” y “ShortErrorMessage”. La intención de este índice es que sea utilizado a la hora de buscar información en el sistema de manera semántica, con entendimiento del contexto por el cual se está buscando, en lugar de simplemente buscar correspondencias exactas entre fragmentos de texto como se puede hacer con las consultas normales.

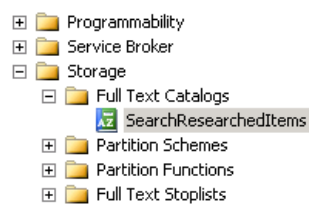


Imagen 1: Catálogo de texto creado para habilitar la búsqueda semántica.

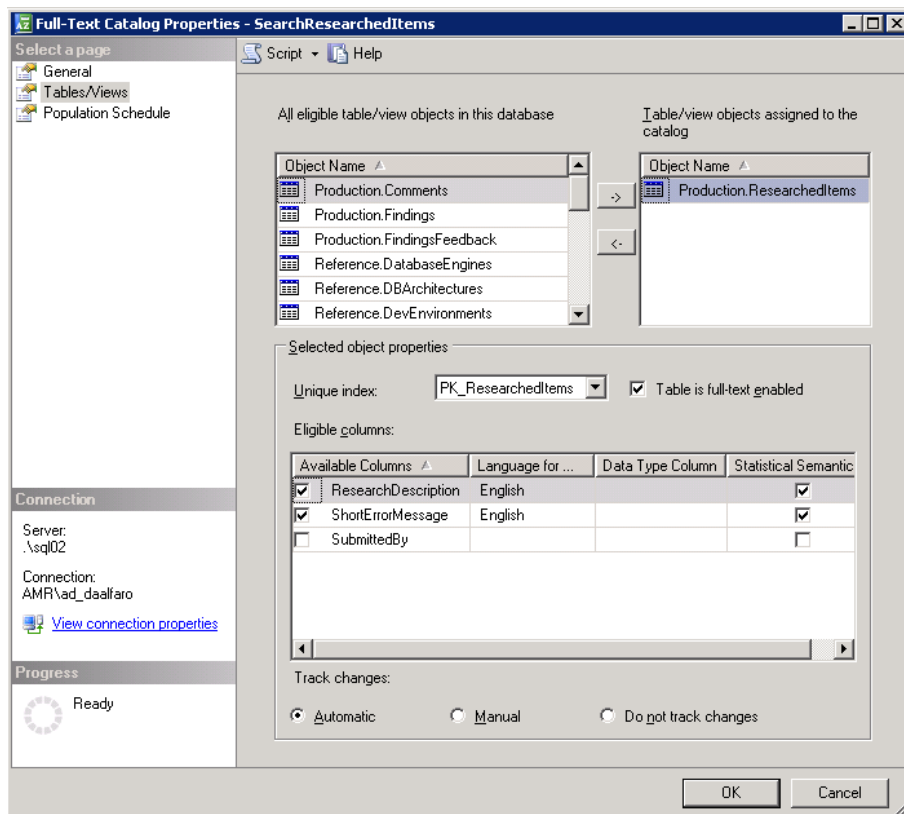


Imagen 2: Configuración del catálogo para habilitar la búsqueda semántica.

Diseño de la Interfaz Gráfica y Usabilidad:

Al momento de decidir cómo iban a interactuar los usuarios con el sistema de gestión de conocimiento, desde el inicio se decidió desarrollar una aplicación web. Entre las razones por las cuales esa decisión se tomó sobresale la facilidad para actualizar el contenido o para poner versiones nuevas en producción sin tener que hacer que los usuarios actualicen un cliente. Al decidir el lenguaje de programación a ser utilizado, se consideraron los lenguajes más comunes en Intel: Java, Visual C# y Visual Basic. Se decidió utilizar C# al ser este el lenguaje con el cual el desarrollador de la aplicación se siente más cómodo en trabajar.

En términos de usabilidad, el principal requisito que se identificó en las diferentes entrevistas con desarrolladores es que fuera fácil y rápido de utilizar. Con esto en mente, se implementó el concepto de "plantillas" dentro de la herramienta. La herramienta pide a los usuarios metadatos de los sistemas o ambientes en los cuales están trabajando, por ejemplo el motor de bases de datos o el lenguaje de programación utilizado. En total se capturan ocho características de cada aplicación o ambiente en el que trabaja el desarrollador. Las plantillas permiten a los usuarios crear sinónimos o asignarle nombres a un conjunto de características específico, de manera que

el proceso largo de escoger las características sólo se hace una vez (por plantilla). Al utilizar la herramienta, el usuario referencia a las plantillas que ha creado en lugar de seleccionar las características manualmente cada vez, con lo cual se facilita mucho la interacción con la herramienta.

En materia de la estética del sistema, al no contar con un recurso dedicado al diseño gráfico, se procedió a visitar el sitio web de Andreas Viklund³. En dicho sitio se pueden encontrar plantillas para sitios web de alta calidad y de manera gratuita.

Diseño del Modelo de Minería de Datos:

En este proyecto, se utiliza la minería de datos para agrupar las plantillas utilizadas por los usuarios de acuerdo a su similaridad. La utilidad de dicha agrupación radica en reducir el conjunto de datos a ser consultado al realizar una búsqueda avanzada por medio de la herramienta. La forma en que esto se consigue, es creando un índice de grupo o clúster en la tabla donde se almacenan las plantillas. Aquí es donde las características recolectadas al momento de elaboración de las plantillas son utilizadas, como entradas para el modelo de minería de datos. Al ingresar una plantilla nueva, se consulta el modelo de minería de datos y este determina el grupo al que pertenece según sus características. El índice obtenido al consultar el modelo se ingresa a la tabla de plantillas junto con el resto de la información.

La imagen a continuación muestra el modelo de minería de datos creado y configurado para utilizar las características de las plantillas como datos de entrada para predecir un número de clúster o grupo.

³ <http://andreasviklund.com/templates/#all/1/list>

Structure	Clustering2
Cluster ID	Microsoft_Clustering
Database Engine	PredictOnly
DB Version	Input
Dev Env Version	Input
Dev Environment	Input
Dev Lang Version	Input
Dev Language	Input
ID Template	Key
OS Technology	Input
OS Version	Input
OS	Input
Version	Input
Virtualization Option	Input
Web Server Version	Input
Web Server	Input

Imagen 3: Modelo de Minería de Datos Utilizado

Al realizar una consulta avanzada por medio de la herramienta, se especifican las características que le interesan al usuario y con estas se vuelve a consultar el modelo. Dicha consulta devuelve un índice en el cual, basado en las características, el modelo dicta que se encuentran aplicaciones o plantillas similares a la cual se está consultando. Este índice se utiliza en un filtro a la hora de buscar ítems según los criterios especificados, y se reduce el conjunto de datos total donde se realiza esta consulta.

Como no se conoce cuántos "tipos" de aplicaciones se van a utilizar, ni se puede determinar a priori, es necesario que el modelo de minería de datos sea capaz de determinar una cantidad apropiada de grupos o clústeres de aplicaciones. Debido esto, no se pueden utilizar algoritmos de clasificación, sino que se pueden utilizar sólo algoritmos de agrupación ("clustering"). Adicionalmente, no se debe configurar el algoritmo utilizado para trabajar con una cantidad determinada de grupos, sino que se debe permitir al algoritmo determinar la cantidad idónea de grupos. Para hacer esto, al algoritmo se le configura el parámetro `CLUSTER_COUNT` con un valor de 0, tal y como muestra la siguiente imagen:

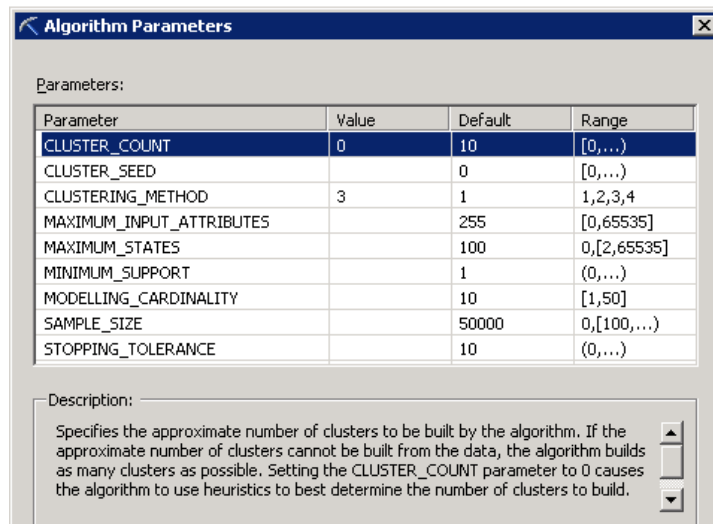


Imagen 4: Configuración del modelo de minería de datos

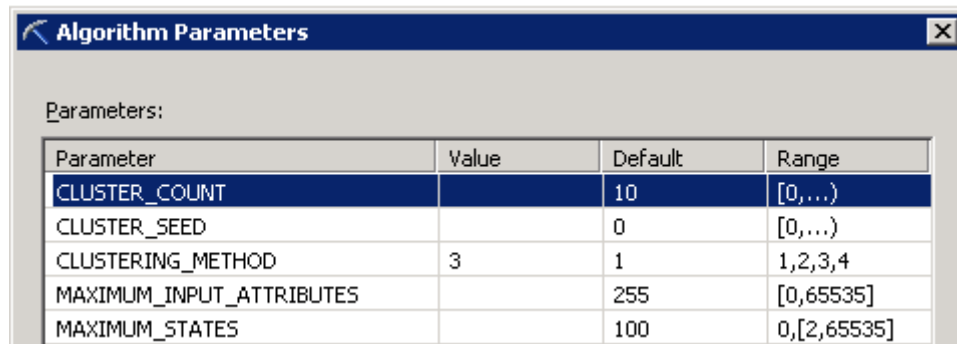
La utilización de este valor tiene implicaciones que pueden parecer adversas para demostrar la utilidad y funcionalidad del sistema propuesto. Al dejar que sea el algoritmo quien determine la cantidad idónea de grupos para trabajar, si no se cuenta con un amplio conjunto de datos, se abre la puerta para que el algoritmo determine que se puede trabajar solo con un grupo. Esto se comprueba al hacer una consulta al modelo de minería de datos. Como se puede apreciar en la siguiente imagen, al realizar una consulta al modelo de minería de datos con este parámetro configurado, todas las plantillas son asignadas al mismo grupo (Clúster 1):

IDTemplate	TemplateVersion	\$CLUSTER
2	1	Cluster 1
3	1	Cluster 1
4	1	Cluster 1
4	2	Cluster 1
4	3	Cluster 1
4	4	Cluster 1
4	5	Cluster 1
4	6	Cluster 1
4	7	Cluster 1
4	8	Cluster 1
4	9	Cluster 1
4	10	Cluster 1
4	11	Cluster 1
4	12	Cluster 1
5	1	Cluster 1
5	2	Cluster 1
6	1	Cluster 1
6	2	Cluster 1

Imagen 5: Resultado de la consulta de prueba al modelo de minería de datos

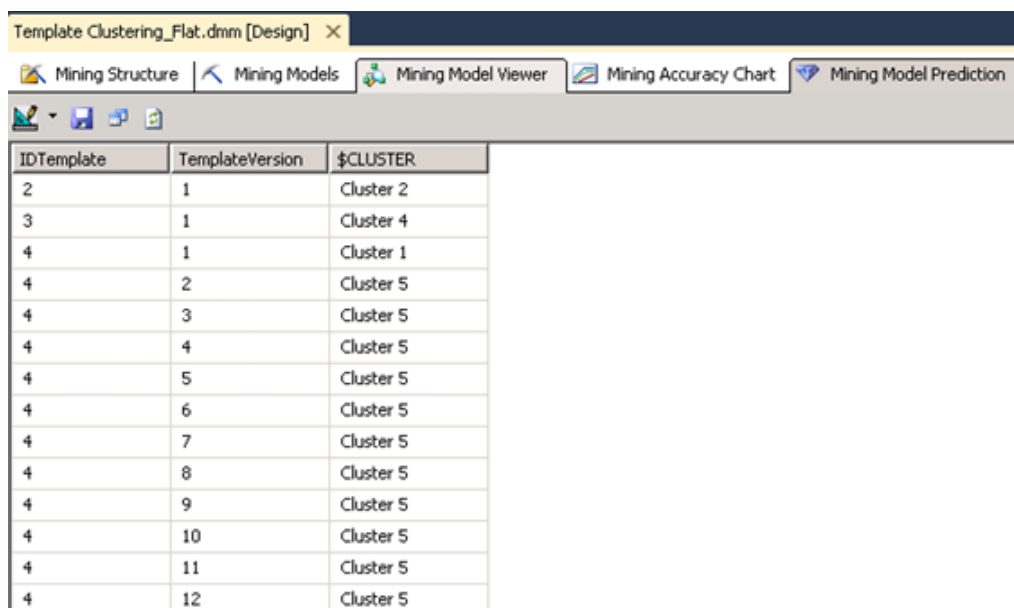
Sin poner una cantidad estática de clústeres.

Para demostrar que la funcionalidad propuesta en este proyecto sí funciona, se puede eliminar el valor ingresado al parámetro CLUSTER_COUNT para utilizar el valor por defecto de la herramienta (10). Al aplicar este cambio y volver a consultar el modelo, se puede apreciar el efecto que trabajar con una cantidad estática tiene sobre la agrupación. Las siguientes imágenes muestran el cambio en la configuración y el resultado de la consulta al modelo de minería de datos.



Parameter	Value	Default	Range
CLUSTER_COUNT		10	[0,...)
CLUSTER_SEED		0	[0,...)
CLUSTERING_METHOD	3	1	1,2,3,4
MAXIMUM_INPUT_ATTRIBUTES		255	[0,65535]
MAXIMUM_STATES		100	0,[2,65535]

Imagen 6: Algoritmo re-configurado para probar que la agrupación por características realmente funciona



IDTemplate	TemplateVersion	\$CLUSTER
2	1	Cluster 2
3	1	Cluster 4
4	1	Cluster 1
4	2	Cluster 5
4	3	Cluster 5
4	4	Cluster 5
4	5	Cluster 5
4	6	Cluster 5
4	7	Cluster 5
4	8	Cluster 5
4	9	Cluster 5
4	10	Cluster 5
4	11	Cluster 5
4	12	Cluster 5

Imagen 5: Resultado de la consulta de prueba al modelo de minería de datos con una cantidad estática de clústeres (10).

Algoritmo de Minería de Datos

SQL Server 2012 ofrece 2 algoritmos diferentes para realizar el “clustering”: EM (Maximización de Expectativas) y KMeans. Para cada uno de estos algoritmos, se ofrecen dos variaciones: Escalable y No Escalable. La diferencia entre estas variaciones radica en que la versión escalable trabaja inicialmente con 50 mil registros y no lee ningún otro registro a menos de que lo(s) necesite para “afinar” el modelo. En contraste, la versión No Escalable trabaja con todo el conjunto de datos. Para este proyecto, se eligió el algoritmo KMeans en su versión escalable. Debido a la cantidad inicial de datos (considerablemente por debajo de 50 mil), y a la naturaleza del conjunto de datos, no es necesario diferenciar entre escalable y no escalable.

Justificación Matemática del Algoritmo

La justificación matemática del algoritmo KMeans requiere la definición de varios conceptos:

- **Centro de Gravedad:** coordenada que sería el centro de la masa (\mathbf{g}) o el centro de una partición o clúster [$\mathbf{g}(k)$]. Sus fórmulas se presentan a continuación:

$$\mathbf{g} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad \left| \quad \mathbf{g}_k = \frac{1}{|C_k|} \sum_{i \in C_k} \mathbf{x}_i$$

- **Inercia:** concepto que se utiliza dentro del contexto de este algoritmo para medir la homogeneidad de las clases o clústeres (intra-clase) y la separación entre los diferentes clústeres (inter-clase).
- **Inercia inter-clase:** inercia de centros de gravedad respecto al centro de gravedad total.

$$B(P) = \sum_{k=1}^K \frac{|C_k|}{n} \|\mathbf{g}_k - \mathbf{g}\|^2$$

- **Inercia intra-clase:** inercia al interior de cada clase.

$$W(P) = \sum_{k=1}^K I(C_k) = \frac{1}{n} \sum_{k=1}^K \sum_{i \in C_k} \|\mathbf{x}_i - \mathbf{g}_k\|^2$$

- **Inercia total:** sumatoria de la inercia intra-clase e inter-clase.

El objetivo del algoritmo es encontrar esas agrupaciones donde la inercia inter-clase sea la máxima y la inercia intra-clase sea mínima. En otras palabras, que los centros de gravedad de cada grupo estén lo más lejanos entre sí posible, y que los puntos de cada clúster esté lo más cerca posible a su respectivo centro de gravedad.

Ejemplo de consulta

El proceso de consulta avanzada, donde se combina la información del modelo de minería de datos con el resultado de una búsqueda semántica, consta de dos partes:

- 1) Primero se consulta el modelo de minería de datos, con el fin de obtener el número de grupo o clúster al cual pertenecería la aplicación con las características especificadas por el usuario.
- 2) Posteriormente, ya conociendo el número de clúster, se procede a hacer la consulta semántica para recuperar aquellos “posts” que contienen palabras similares a las cuales se están buscando. El resultado de esta consulta se filtra utilizando el número de clúster.

A continuación se presenta una consulta hecha por medio de la interfaz y la correspondiente sentencia capturada utilizando el SQL Profiler:

Search Posts

No rows returned. Please try again.

Basic Search
 Advanced Search
 Specific Post

Error Message:

Description or keywords:

Database Engine:	<input type="text" value="Not Applicable"/> <input type="text" value="MS SQL Server 2005"/> <input type="text" value="SQL Server 2008"/>	Development Environment:	<input type="text" value="Not Applicable"/> <input type="text" value="MS Visual Studio 2010 SP1"/>
Development Language:	<input type="text" value="Not Applicable"/> <input type="text" value="Visual C# 2010"/>	Host Server OS:	<input type="text" value="Not Applicable"/> <input type="text" value="Windows Server 2008 R2 SP3 (64-bit)"/> <input type="text" value="Windows Server 2008 R2 SP2 (64-bit)"/>
Virtualization Options:	<input type="text" value="Not Applicable"/> <input type="text" value="Hyper-V"/>	Web Server Technology:	<input type="text" value="Not Applicable ()"/> <input type="text" value="IIS (6.0)"/> <input type="text" value="IIS (7.0)"/>

Imagen 6: consulta avanzada realizada por medio de la interfaz.

Código capturado:

```

SELECT FLATTENED PredictNodeId([Clustering2].[Cluster ID])
From [Clustering2]
PREDICTION JOIN
  (SELECT 'MS SQL Server' AS [DatabaseEngine], '2005' AS DBVersion, 'Not Applicable' AS
  [DevEnvironment], '' AS [DevEnvVersion], 'Visual C#' AS [DevLanguage], '2010' AS
  [DevLangVersion], 'Windows' AS [OS], 'Server 2008 R2 SP3' AS [OSVersion], '64-bit' AS
  [OSTechnology], '1' AS [VirtualizationOption], 'IIS' AS [webServer], '6.0' AS [WebServerVersion])
  AS t
ON [Clustering2].[Database Engine] = t.[DatabaseEngine]
  
```

```

AND [Clustering2].[DB Version] = t.[DBVersion]
AND [Clustering2].[Dev Environment] = t.[DevEnvironment]
AND [Clustering2].[Dev Env Version] = t.[DevEnvVersion]
AND [Clustering2].[Dev Language] = t.[DevLanguage]
AND [Clustering2].[Dev Lang Version] = t.[DevLangVersion]
AND [Clustering2].[OS] = t.[OS]
AND [Clustering2].[OS Version] = t.[OSVersion]
AND [Clustering2].[OS Technology] = t.[OSTechnology]
AND [Clustering2].[Virtualization Option] = t.[VirtualizationOption]
AND [Clustering2].[Web Server] = t.[WebServer]
AND [Clustering2].[Web Server Version] = t.[WebServerVersion]

```

El resultado de esta consulta se utiliza como filtro en el procedimiento almacenado de la base de datos, en el cual se realiza la búsqueda semántica, tal y como lo muestra la siguiente imagen:

```

select RI.IDResearchedItem, RI.ShortErrorMessage, RI.ResearchDescription, RI.DateAdded, RI.SubmittedBy , FT.[Rank]
from FREETEXTTABLE
    ([Production].[ResearchedItems],
     [ResearchDescription],
     @Description) FT
inner join [Production].[ResearchedItems] RI on RI.IDResearchedItem = ft.[key]
inner join [Reference].[Templates] T on (T.IDTemplate = ri.IDTemplate and t.Version = ri.TemplateVersion)
Where T.ClusterID = @ClusterID

```

Imagen 7: fragmento de código que realiza la consulta semántica.

Análisis y Discusión de Resultados

Los resultados obtenidos al final de la etapa de desarrollo de la solución son bastante satisfactorios. Se cumplieron los objetivos que se propusieron para el proyecto y se logró desarrollar un sistema que cumple no solo con las funciones necesarias para compartir conocimiento a través de la compañía, sino que también cumple con las características de facilidad de uso y agilidad que los usuarios deseaban tener.

La utilización de las plantillas facilita a los usuarios ingresar los datos básicos de sus herramientas. A la vez, permiten que los usuarios puedan utilizarlas como atajos para no tener que llenar la información cada vez que quieran compartir algo. Esta agilidad para interactuar con la herramienta busca motivar a los desarrolladores a favorecer la utilización de esta herramienta ante otras posibles opciones internas o externas.

A nivel de utilidad, el haber implementado la búsqueda semántica es otro punto a favor de la herramienta. Con esta ventaja, la herramienta puede competir con otras soluciones como sitios de SharePoint o foros disponibles en internet. Claro está que el sistema desarrollado no es tan robusto ni tiene un ámbito tan grande como los buscadores que hay disponibles en internet, pero el proveerle una búsqueda más avanzada y comprensiva que una búsqueda que busque cadenas de texto exactos le permite ser un competidor digno de atención. Aunado a esto, el poder complementar la búsqueda semántica mejora aún más la eficiencia del sistema en cuanto a búsquedas se refiere. Esta mejoría no solo se vería en términos de velocidad de respuesta, pues el contar con un índice de similaridad permite filtrar muchos de los registros de la base de datos, sino que también en la calidad de los resultados mostrados al usuario. El utilizar este índice de similaridad para buscar artículos o ítems que se hayan ingresado con características similares a las que busca el usuario de turno, permite presentar al usuario resultados más significativos para su búsqueda.

Adicionalmente, gracias a las opciones con las que cuenta el motor de bases de datos para automatizar tareas, se facilita en gran manera el mantenimiento de la herramienta. Se cuenta con tareas automatizadas (“Jobs”) para mantener actualizado el modelo de minería de datos: éste se procesa cada hora con la totalidad del conjunto de datos, lo cual garantiza que al ingresar una plantilla nueva y consultar el modelo para determinar a cuál grupo pertenece, el resultado está basado en un análisis con datos recientes.

Finalmente, la herramienta como un todo cumple con la misión de ofrecer un vehículo para mejorar la colaboración entre los programadores que laboran en Intel. Es una herramienta accesible para todos los empleados (es un sitio web), abierto a todos (no pide autenticación, utiliza la de Windows) y que permite la interacción entre los usuarios (los artículos ingresados pueden ser comentados y votados para determinar su utilidad).

Plan de Inversión del Proyecto

A continuación se presentan dos estimaciones del costo requerido para llevar a cabo el proyecto. La diferencia entre ambos es el costo de las licencias. Cabe destacar que, al tener Intel un convenio corporativo con Microsoft, no es necesario invertir en software pues ya se tiene disponible. De manera similar, la compañía cuenta con una arquitectura de virtualización que permite a los empleados disponer de máquinas virtuales para proyectos específicos sin costos directos para el departamento al que el empleado pertenece.

Como referencia, se incluye el cálculo de los costos si este proyecto fuera a ser llevado a cabo en una compañía diferente a Intel.

Estimación 1:

	Costo Fuera de Intel	Costo para Intel
Tiempo del Administrador de Proyecto	\$ 15.686,27	\$ 15.686,27
Tiempo del Desarrollador	\$ 1.176,47	\$ 1.176,47
Licencia SQL Server 2008 R2 Enterprise Edition ⁴	\$ 28.749,00	\$ -
Licencia Visual Studio 2010 Ultimate ⁵	\$ 11.899,00	\$ -
Licencia Windows Server 2008 R2 Enterprise ⁶	\$ 3.919,00	\$ -
Servidor o Máquina Virtual ⁷	\$ 4.939,00	\$ -
Total	\$ 66.368,75	\$ 16.862,75

Estimación 2:

	Costo Fuera de Intel	Costo para Intel
Tiempo del Administrador de Proyecto	\$ 15.686,27	\$ 15.686,27
Tiempo del Desarrollador	\$ 1.176,47	\$ 1.176,47
Licencia SQL Server 2008 R2 Standard Edition	\$ 7.499,00	N/A
Licencia Windows Server 2008 R2 Standard	\$ 1.209,00	N/A
Licencia Visual Studio 2010 Professional ⁸	\$ 590,00	N/A
Servidor o Máquina Virtual	\$ 4.939,00	\$ -
Total	\$ 31.099,75	\$ 16.862,75

⁴ <http://www.brentozar.com/archive/2009/11/sql-server-2008-r2-pricing-and-feature-changes/>

⁵ http://www.amazon.com/Visual-Studio-2010-Ultimate-MSDN/dp/B0038KNER0/ref=sr_1_2?s=software&ie=UTF8&qid=1333840318&sr=1-2

⁶ <http://news.softpedia.com/news/Windows-Server-2008-R2-Pricing-and-Licensing-134153.shtml>

⁷ <http://www.dell.com/cr/empresas/p/poweredge-r720xd/pd>

⁸ http://www.amazon.com/Microsoft-C5E-00521-Visual-Studio-Professional/dp/B0038KTO8S/ref=sr_1_3?s=software&ie=UTF8&qid=1333839437&sr=1-3

Conclusiones y Trabajos Futuros

Luego del desarrollo satisfactorio de la solución para compartir conocimiento entre desarrolladores, queda un sentimiento de satisfacción por haber podido cumplir con las metas propuestas. Asimismo, queda la satisfacción de haber podido integrar en una misma solución dos áreas muy importantes en la actualidad como lo son la minería de texto y la minería de datos estructurados. Finalmente, también cabe resaltar la satisfacción de haber podido identificar un área que puede ser mejorada dentro de una compañía tan grande como Intel, como lo es la colaboración entre programadores, y de proponer una solución integral a dicha necesidad.

Sin embargo, aún con el trabajo expuesto en este documento, quedan todavía muchas oportunidades que se pueden desarrollar con base en este proyecto, así como tareas pendientes de este mismo proyecto. Entre las tareas pendientes, la tarea más fuerte es abogar por la adaptación de una herramienta de esta naturaleza a un proceso tan crítico para la empresa como lo es el desarrollo de, y el soporte al, software. Esta herramienta fue desarrollada sin que haya sido identificada como una necesidad crítica para alguna organización. Más bien, se presenta como una herramienta para satisfacer una necesidad que muchas organizaciones tienen y no tienen el tiempo o el presupuesto para dedicar recursos a solucionarla, como lo es la transferencia de conocimientos entre programadores. Debido a esto, el primer paso en las etapas posteriores al desarrollo es la promoción de la herramienta, ofreciéndola como una solución integral que busca mejorar la colaboración entre programadores y/o personal de soporte.

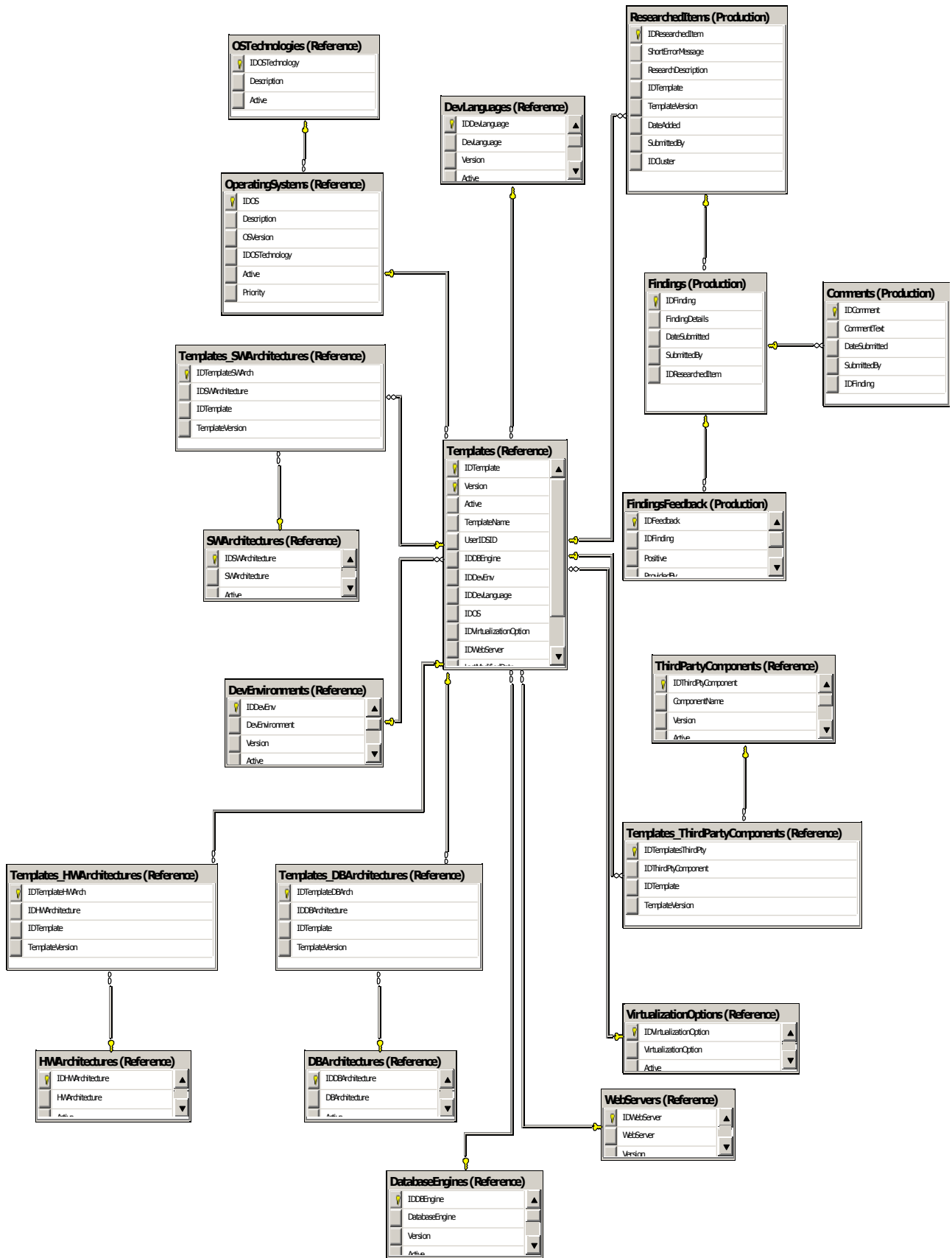
Aunado a lo anterior, la aceptación y propagación de la herramienta, es necesario seguir con el flujo de proyecto que propone la metodología Lean Six Sigma. El proyecto se presenta al salir de la etapa de *Implementación*, siendo esta la cuarta de cinco etapas. La última etapa corresponde a la etapa de *Control*, donde se volverán a visitar las métricas definidas (horas dedicadas a buscar soluciones y cantidad de sitios visitados) para determinar el nivel de éxito de este proyecto. Sin embargo, antes de poder llegar a la etapa de control, es necesario que la herramienta sea adoptada oficialmente y utilizada.

El trabajo desarrollado abre las puertas para seguir trabajando en proyectos de esta naturaleza. Con base en lo aprendido durante este proyecto, se considera proponer el desarrollo de un sistema de administración de conocimiento que se catalogue automáticamente. Dicha propuesta buscaría explotar aún más las ventajas que ofrecen los índices de texto (Full-Text

Indexes) y la automatización de tareas, sin dejar atrás la parte de minería de datos. En esta nueva etapa, se buscaría poder automatizar la creación de un catálogo de palabras clave en columnas de texto no estructurado, que puedan ser eventualmente analizadas y agrupadas de manera similar a como se agruparon los datos en este proyecto. De este esfuerzo podrían salir nuevas formas de entender los datos no estructurados que se almacenan en muchas de las bases de datos de la compañía, y podría iniciarse una tendencia hacia el análisis de datos no estructurados sin la necesidad de trabajar con sistemas más elaborados como Hadoop, aptos para proyectos de mayor escala.

Anexos

Diagrama Entidad - Relación



Páginas del Sistema

Página de Inicio:

The screenshot shows the homepage of 'The Developer's KMDB' in a Firefox browser window. The browser's address bar shows 'vmsfstrng101.81'. The page title is 'The Developer's KMDB' with the subtitle 'Promoting developer collaboration across Intel'. The main content area features a large blue header image. Below the header, there are several sections: 'Home', 'Manage Templates', 'Template Metadata', 'Add Research Item', 'Search Posts', and 'My Posts' on the left sidebar. The main content area includes a 'Welcome to The Developer's KMDB' section, followed by 'Why use it?', 'How do I use it?', and a list of actions: 'Adding Research Items', 'Provide or Complement Findings', and 'Need to Find an Expert?'. The browser's status bar at the bottom shows 'vmsfstrng101.81/default.aspx sts'.

This screenshot shows the same 'The Developer's KMDB' homepage, but with more content visible. The browser's address bar shows 'vmsfstrng101.81/default.aspx'. The left sidebar now includes 'Latest Posts' and 'Hall of Fame' sections. The main content area includes a 'Where Should I Start?' section. The 'Adding Research Items' section is expanded to include 'Provide or Complement Findings' and 'Need to Find an Expert?'. The 'How do I use it?' section is also expanded to include 'Adding Research Items', 'Provide or Complement Findings', 'Search Existing Posts', 'Provide Feedback', and 'Need to Find an Expert?'. The browser's status bar at the bottom shows 'vmsfstrng101.81/default.aspx sts'. The footer of the page indicates 'Logged in as daalfaro' and 'Template design by Andreas Viktund'.

Página de Ingreso de Plantillas:

Firefox

Template Management

vmsfstrng101.81/ManageTemplates.aspx

Most Visited Getting Started Home - L2 DSS Hub Suggested Sites Web Slice Gallery Welcome to DSS Hub ...

Home

Manage Templates

Create or edit your application templates.

Template Metadata

Add new entries to our metadata catalogues to be used in your templates.

Add Research Item

Share your knowledge or seek help!

Search Posts

Search for other users' posts based on application characteristics.

My Posts

See posts you have created plus all comments.

Latest Posts

Checkout the most recent posts that have been submitted.

Hall of Fame

See our top contributors!

Manage Your Templates

What would you like to do? Add New Template Change Template

NOTE: If you can't find an element that doesn't fit your need, please select 'Not Applicable' or visit our Template Metadata section to add it.

Database Engine:
 Not Applicable
 MS SQL Server 2005 RTM
 MS SQL Server 2005 SP3
 MS SQL Server 2008 RTM

Development Language:
 Not Applicable
 Java
 Visual Basic 2010
 Visual C# 2010

Virtualization Options:
 Not Applicable
 Hyper-V
 VMWare

DB architectures used:
 Not Applicable
 Mirroring
 RAC
 Replication

HW Architectures used:
 Not Applicable
 Cluster

Template Name:

Development Environment:
 Not Applicable
 JDK 7
 Visual Studio 2010 Ultimate

Host Server OS:
 Not Applicable
 MS Windows 2003 (64-bit)
 MS Windows 2003 R2 (64-bit)
 MS Windows 2008 (64-bit)

Web Server Technology:
 Not Applicable ()
 Apache ()
 IIS (6.0)
 IIS (7.0)

SW architectures used:

Third party components used:
 Not Applicable ()
 Infragistics (2011)
 Infragistics (2012)

Logged in as daalfaro
 Template design by Andreas Viklund

Firefox

Template Management

vmsfstrng101.81/ManageTemplates.aspx

Most Visited Getting Started Home - L2 DSS Hub Suggested Sites Web Slice Gallery Welcome to DSS Hub ...

Home

Manage Templates

Create or edit your application templates.

Template Metadata

Add new entries to our metadata catalogues to be used in your templates.

Add Research Item

Share your knowledge or seek help!

Search Posts

Search for other users' posts based on application characteristics.

My Posts

See posts you have created plus all comments.

Latest Posts

Checkout the most recent posts that have been submitted.

Hall of Fame

See our top contributors!

Manage Your Templates

What would you like to do? Add New Template Change Template

NOTE: If you can't find an element that doesn't fit your need, please select 'Not Applicable' or visit our Template Metadata section to add it.

Choose a Template:
 Dev KMDB

Database Engine:
 MS SQL Server 2008 RTM
 MS SQL Server 2008 SP3
 MS SQL Server 2012 RTM
 MS SQL Server 2012 SP1

Development Language:
 Not Applicable
 Java
 Visual Basic 2010
 Visual C# 2010

Virtualization Options:
 Not Applicable
 Hyper-V
 VMWare

DB architectures used:
 Not Applicable
 Mirroring
 RAC
 Replication

HW Architectures used:
 Not Applicable
 Cluster

Template Name:
 Dev KMDB

Development Environment:
 Not Applicable
 JDK 7
 Visual Studio 2010 Ultimate

Host Server OS:
 MS Windows 2003 (64-bit)
 MS Windows 2003 R2 (64-bit)
 MS Windows 2008 (64-bit)
 MS Windows 2008 R2 (64-bit)

Web Server Technology:
 Not Applicable ()
 Apache ()
 IIS (6.0)
 IIS (7.0)

SW architectures used:

Third party components used:
 Not Applicable ()
 Infragistics (2011)
 Infragistics (2012)

Logged in as daalfaro
 Template design by Andreas Viklund

Administración de Metadatos de las Plantillas:

Firefox

http://vmsfstrng10...plateMetadata.aspx

vmsfstrng101:81/TemplateMetadata.aspx

Most Visited Getting Started Home - L2 DSS Hub Suggested Sites Web Slice Gallery Welcome to DSS Hub ...

Home

Template Metadata Management

What would you like to do? Add What kind of item? DB Engine

Name: Version:

Clear Add

Manage Templates
Create or edit your application templates.

Template Metadata
Add new entries to our metadata catalogues to be used in your templates.

Add Research Item
Share your knowledge or seek help!

Search Posts
Search for other users' posts based on application characteristics.

My Posts
See posts you have created plus all comments.

Latest Posts
Checkout the most recent posts that have been submitted.

Hall of Fame
See our top contributors!

Firefox

http://vmsfstrng10...plateMetadata.aspx

vmsfstrng101:81/TemplateMetadata.aspx

Most Visited Getting Started Home - L2 DSS Hub Suggested Sites Web Slice Gallery Welcome to DSS Hub ...

Home

The Developer's KMDB

Promoting developer collaboration across Intel

Template Metadata Management

What would you like to do? Remove What kind of item? DB Engine

Choose the item(s) to remove:

Clear Remove

Not Applicable
MS SQL Server 2005 RTM
MS SQL Server 2005 SP3
MS SQL Server 2008 RTM

NOTE: Item deletions are permanent and can't be undone.

Manage Templates
Create or edit your application templates.

Template Metadata
Add new entries to our metadata catalogues to be used in your templates.

Add Research Item
Share your knowledge or seek help!

Search Posts
Search for other users' posts based on application characteristics.

My Posts
See posts you have created plus all comments.

Latest Posts

Página para agregar un "Research Item" nuevo:

Firefox

Add New Research Item

vmsfistmg101:81/AddResearchItem.aspx

Most Visited Getting Started Home - L2 DSS Hub Suggested Sites Web Slice Gallery Welcome to DSS Hub ...

The Developer's KMDB

Promoting developer collaboration across Intel

Add New Research Item

NOTE: Please make sure that you have already created an application template. If you haven't, please visit our 'Manage Templates' section

daalfaro

Choose an application template: Skills Analyzer

What error message did you get? (if applicable)

Full Description

have a solution for this item

Add Clear

vmsfistmg101:81/AddResearchItem.aspx

Firefox

Add New Research Item

vmsfistmg101:81/AddResearchItem.aspx

Most Visited Getting Started Home - L2 DSS Hub Suggested Sites Web Slice Gallery Welcome to DSS Hub ...

The Developer's KMDB

Promoting developer collaboration across Intel

Add New Research Item

NOTE: Please make sure that you have already created an application template. If you haven't, please visit our 'Manage Templates' section

daalfaro

Choose an application template: Skills Analyzer

What error message did you get? (if applicable)

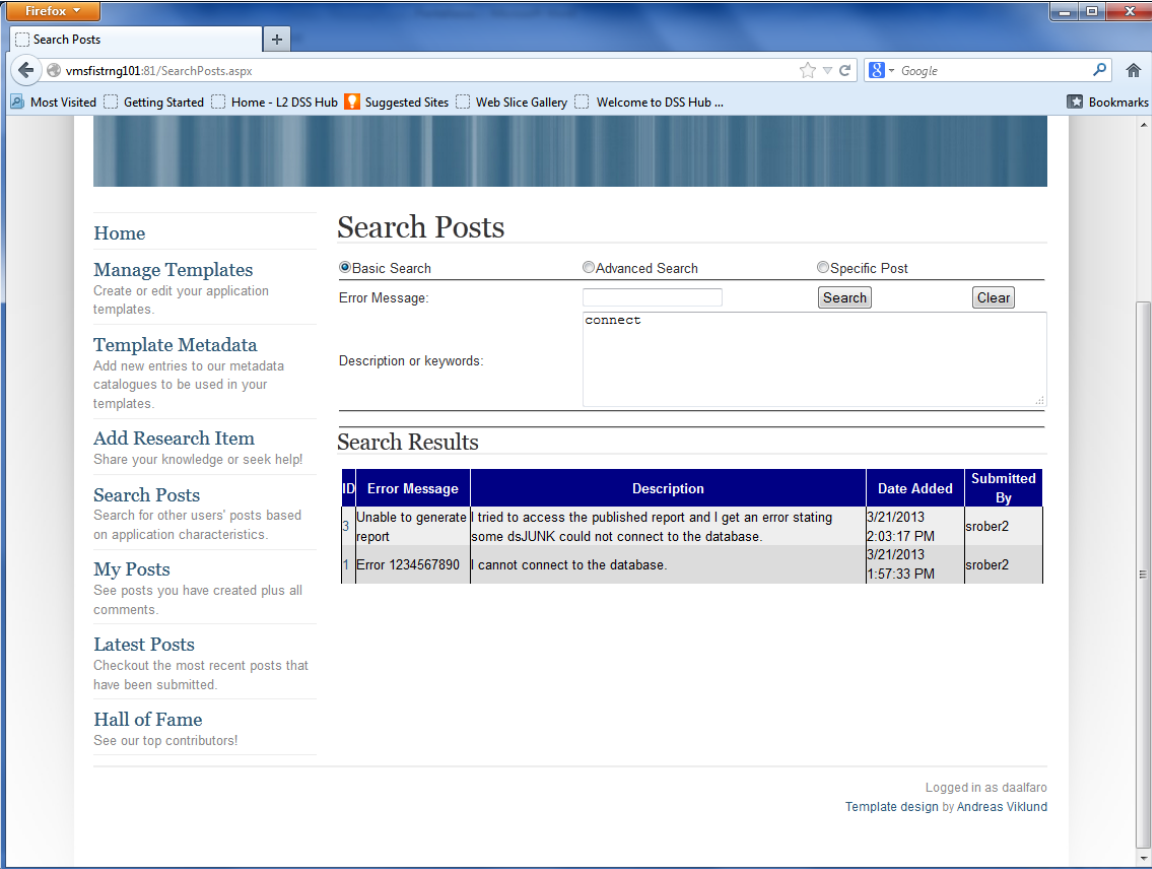
Full Description

have a solution for this item

Add Clear

Logged in as daalfaro
Template design by Andreas Viklund

Página para buscar "posts":



Firefox

Search Posts

vmsfistrng101.81/SearchPosts.aspx

Most Visited Getting Started Home - L2 DSS Hub Suggested Sites Web Slice Gallery Welcome to DSS Hub ...

Home

Manage Templates

Create or edit your application templates.

Template Metadata

Add new entries to our metadata catalogues to be used in your templates.

Add Research Item

Share your knowledge or seek help!

Search Posts

Search for other users' posts based on application characteristics.

My Posts

See posts you have created plus all comments.

Latest Posts

Checkout the most recent posts that have been submitted.

Hall of Fame

See our top contributors!

Search Posts

Basic Search Advanced Search Specific Post

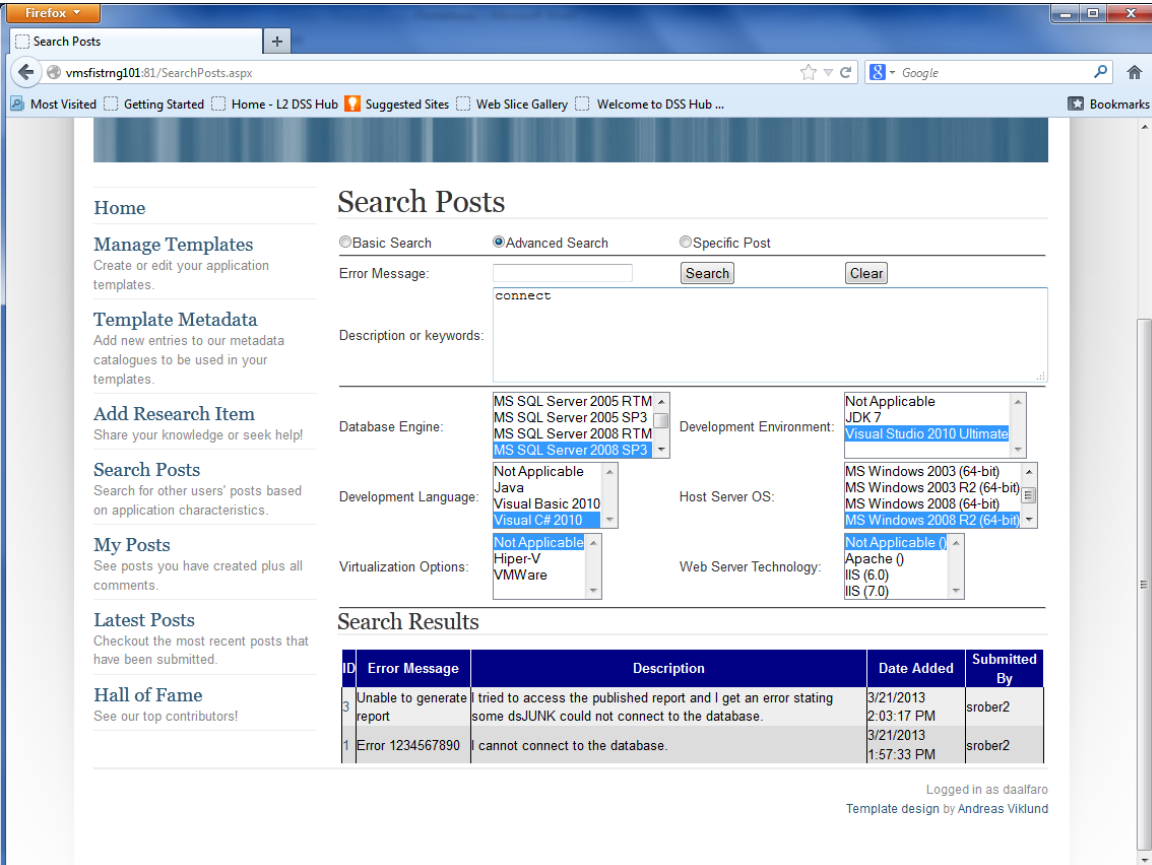
Error Message: Search Clear

Description or keywords:

Search Results

ID	Error Message	Description	Date Added	Submitted By
3	Unable to generate report	tried to access the published report and I get an error stating some dsJUNK could not connect to the database.	3/21/2013 2:03:17 PM	srober2
1	Error 1234567890	cannot connect to the database.	3/21/2013 1:57:33 PM	srober2

Logged in as daalfaro
Template design by Andreas Viklund



Firefox

Search Posts

vmsfistrng101.81/SearchPosts.aspx

Most Visited Getting Started Home - L2 DSS Hub Suggested Sites Web Slice Gallery Welcome to DSS Hub ...

Home

Manage Templates

Create or edit your application templates.

Template Metadata

Add new entries to our metadata catalogues to be used in your templates.

Add Research Item

Share your knowledge or seek help!

Search Posts

Search for other users' posts based on application characteristics.

My Posts

See posts you have created plus all comments.

Latest Posts

Checkout the most recent posts that have been submitted.

Hall of Fame

See our top contributors!

Search Posts

Basic Search **Advanced Search** Specific Post

Error Message: Search Clear

Description or keywords:

Database Engine: MS SQL Server 2005 RTM MS SQL Server 2005 SP3 MS SQL Server 2008 RTM MS SQL Server 2008 SP3

Development Environment: Not Applicable JDK 7 Visual Studio 2010 Ultimate

Development Language: Not Applicable Java Visual Basic 2010 Visual C# 2010

Host Server OS: MS Windows 2003 (64-bit) MS Windows 2003 R2 (64-bit) MS Windows 2008 (64-bit) MS Windows 2008 R2 (64-bit)

Virtualization Options: Not Applicable Hiper-V VMWare

Web Server Technology: Not Applicable Apache () IIS (6.0) IIS (7.0)

Search Results

ID	Error Message	Description	Date Added	Submitted By
3	Unable to generate report	tried to access the published report and I get an error stating some dsJUNK could not connect to the database.	3/21/2013 2:03:17 PM	srober2
1	Error 1234567890	cannot connect to the database.	3/21/2013 1:57:33 PM	srober2

Logged in as daalfaro
Template design by Andreas Viklund

Search Posts

Home

Manage Templates
Create or edit your application templates.

Template Metadata
Add new entries to our metadata catalogues to be used in your templates.

Add Research Item
Share your knowledge or seek help!

Search Posts
Search for other users' posts based on application characteristics.

My Posts
See posts you have created plus all comments.

Latest Posts
Checkout the most recent posts that have been submitted.

Hall of Fame
See our top contributors!

Search Posts

Basic Search Advanced Search **Specific Post**

Post ID:

Search Clear

Search Results

ID	Error Message	Description	Date Added	Submitted By
4	HTTP 404	clicked the report link and I get the following error: HTTP 404. I have a presentation due in 20 minutes... PLEASE HELP!!!!	3/21/2013 2:06:39 PM	srober2

Página que muestra los "posts" del usuario actual:

My Posts

The Developer's KMDB

Promoting developer collaboration across Intel

Home

Manage Templates
Create or edit your application templates.

Template Metadata
Add new entries to our metadata catalogues to be used in your templates.

Add Research Item
Share your knowledge or seek help!

Search Posts
Search for other users' posts based on application characteristics.

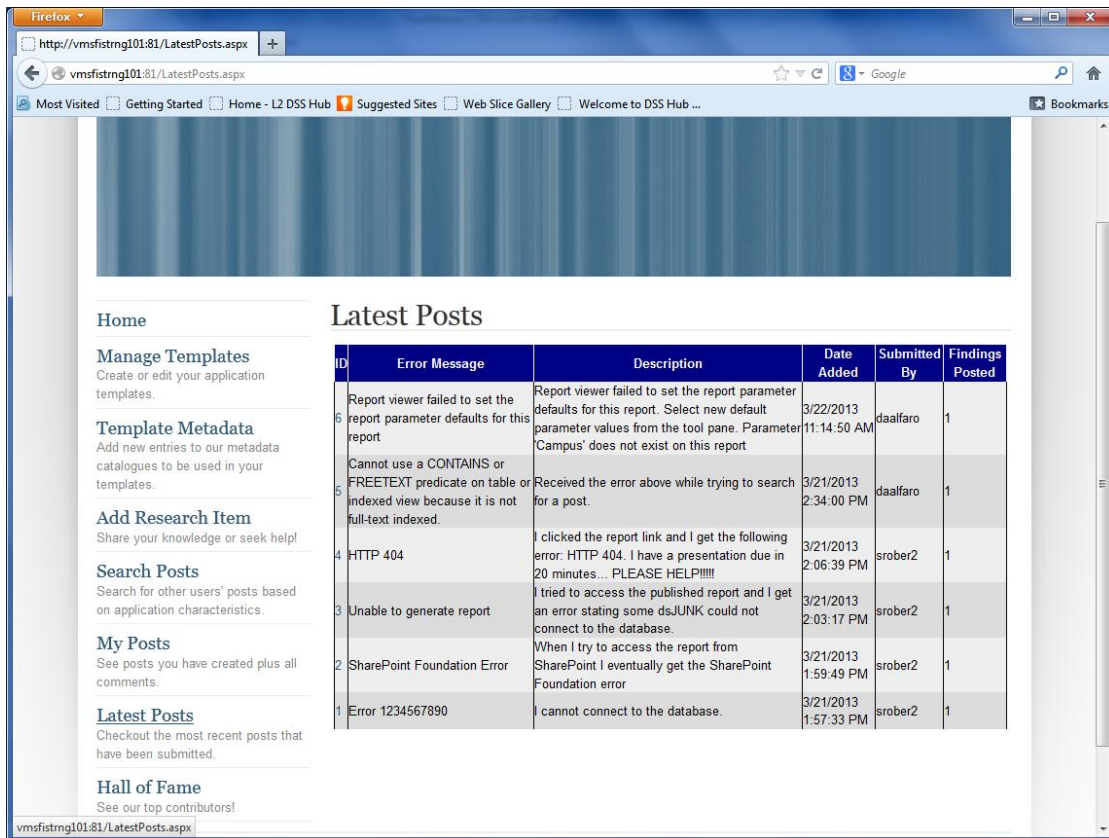
My Posts
See posts you have created plus all comments.

My Posts

ID	Error Message	Description	Date Added	Template	Findings Posted
6	Report viewer failed to set the report parameter d	Report viewer failed to set the report parameter defaults for this report. Select new default	3/22/2013 11:14:50 AM	sCore	1
5	Cannot use a CONTAINS or FREETEXT predicate on tab	Received the error above while trying to search for a post.	3/21/2013 2:34:00 PM	Dev KMDB	1

vmsfistmg101:81/AddResearchItem.aspx

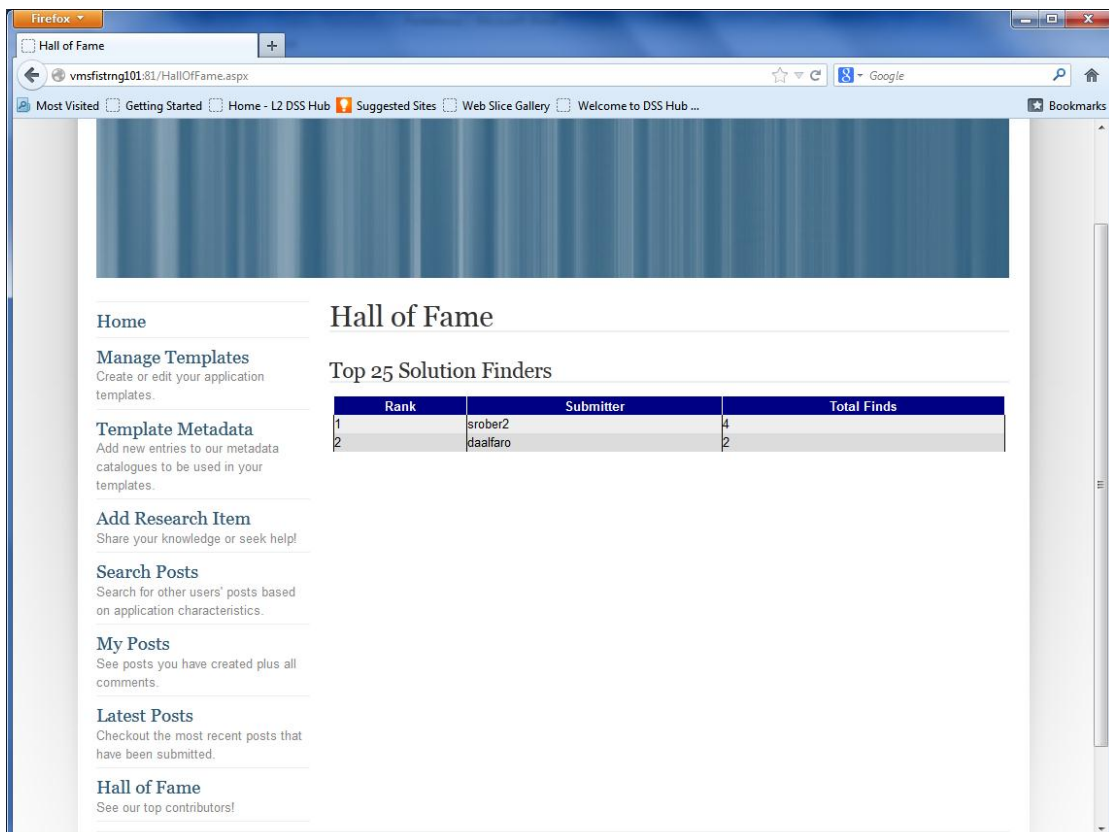
Página que muestra los “posts” más recientes:



The screenshot shows a Firefox browser window displaying the 'Latest Posts' page. The page features a navigation menu on the left with links to Home, Manage Templates, Template Metadata, Add Research Item, Search Posts, My Posts, Latest Posts, and Hall of Fame. The main content area is titled 'Latest Posts' and contains a table of error messages.

ID	Error Message	Description	Date Added	Submitted By	Findings Posted
6	Report viewer failed to set the report parameter defaults for this report	Report viewer failed to set the report parameter defaults for this report. Select new default parameter values from the tool pane. Parameter 'Campus' does not exist on this report	3/22/2013 11:14:50 AM	daalfaro	1
5	Cannot use a CONTAINS or FREETEXT predicate on table or indexed view because it is not full-text indexed.	Received the error above while trying to search for a post.	3/21/2013 2:34:00 PM	daalfaro	1
4	HTTP 404	I clicked the report link and I get the following error: HTTP 404. I have a presentation due in 20 minutes... PLEASE HELP!!!!	3/21/2013 2:06:39 PM	srober2	1
3	Unable to generate report	I tried to access the published report and I get an error stating some dsJUNK could not connect to the database.	3/21/2013 2:03:17 PM	srober2	1
2	SharePoint Foundation Error	When I try to access the report from SharePoint I eventually get the SharePoint Foundation error	3/21/2013 1:59:49 PM	srober2	1
1	Error 1234567890	I cannot connect to the database.	3/21/2013 1:57:33 PM	srober2	1

Página que muestra los usuarios con más hallazgos:



The screenshot shows a Firefox browser window displaying the 'Hall of Fame' page. The page features a navigation menu on the left with links to Home, Manage Templates, Template Metadata, Add Research Item, Search Posts, My Posts, Latest Posts, and Hall of Fame. The main content area is titled 'Hall of Fame' and contains a table of top solution finders.

Rank	Submitter	Total Finds
1	srober2	4
2	daalfaro	2

Diccionario de Datos

Tablas

Esquema: Production

Nombre: Comments

Descripción: Almacena los comentarios ingresados por los usuarios para los diferentes hallazgos.

Columnas:

Columna	Tipo	Longitud	Precisión	Escala
CommentText	varchar	-1	0	0
DateSubmitted	smalldatetime	4	16	0
IDComment	bigint	8	19	0
IDFinding	bigint	8	19	0
SubmittedBy	varchar	100	0	0

Esquema: Production

Nombre: Findings

Descripción: Contiene los hallazgos ingresados por los usuarios.

Columnas:

Columna	Tipo	Longitud	Precisión	Escala
DateSubmitted	datetime	8	23	3
FindingDetails	varchar	-1	0	0
IDFinding	bigint	8	19	0
IDResearchedItem	int	4	10	0
SubmittedBy	varchar	100	0	0

Esquema: Production

Nombre: FindingsFeedback

Descripción: Contiene la retroalimentación dada por los usuarios de la herramienta a los hallazgos.

Columnas:

Columna	Tipo	Longitud	Precisión	Escala
IDFeedback	bigint	8	19	0
IDFinding	bigint	8	19	0
Positive	bit	1	1	0
ProvidedBy	varchar	100	0	0

Esquema: Production

Nombre: ResearchedItems

Descripción: Almacena los diferentes ítems ingresados por los usuarios para cada una de sus plantillas.

Columnas:

DateAdded	datetime	8	23	3
-----------	----------	---	----	---

IDCluster	smallint	2	5	0
IDResearchedItem	int	4	10	0
IDTemplate	int	4	10	0
ResearchDescription	varchar	-1	0	0
ShortErrorMessage	varchar	150	0	0
SubmittedBy	varchar	100	0	0
TemplateVersion	int	4	10	0

Esquema: Reference

Nombre: DatabaseEngines

Descripción: Tabla de referencia que contiene los diferentes motores de bases de datos que pueden ser utilizados por las plantillas que crea el usuario.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
Active	bit	1	1	0
DatabaseEngine	varchar	100	0	0
IDDBEngine	smallint	2	5	0
Priority	tinyint	1	3	0
Version	varchar	100	0	0

Esquema: Reference

Nombre: DBArchitectures

Descripción: Tabla de referencia que contiene las diferentes arquitecturas a nivel de bases de datos que pueden ser utilizados por las plantillas que crea el usuario.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
Active	bit	1	1	0
DBArchitecture	varchar	100	0	0
IDDBArchitecture	int	4	10	0
Priority	tinyint	1	3	0

Esquema: Reference

Nombre: DevEnvironments

Descripción: Tabla de referencia que contiene los diferentes ambientes de desarrollo que pueden ser asociados a las plantillas que crea el usuario.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
Active	bit	1	1	0
DevEnvironment	varchar	150	0	0
IDDevEnv	smallint	2	5	0
Priority	tinyint	1	3	0
Version	varchar	50	0	0

Esquema: Reference

Nombre: DevLanguages

Descripción: Tabla de referencia que contiene los diferentes lenguajes de programación que pueden ser asociados a las plantillas que crea el usuario.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
Active	bit	1	1	0
DevLanguage	varchar	100	0	0
IDDevLanguage	smallint	2	5	0
Priority	tinyint	1	3	0
Version	varchar	50	0	0

Esquema: Reference

Nombre: HWArchitectures

Descripción: Tabla de referencia que contiene las diferentes arquitecturas a nivel de hardware que pueden ser utilizados por las plantillas que crea el usuario.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
Active	bit	1	1	0
HWArchitecture	varchar	100	0	0
IDHWArchitecture	int	4	10	0
Priority	tinyint	1	3	0

Esquema: Reference

Nombre: OperatingSystems

Descripción: Tabla de referencia que contiene los diferentes sistemas operativos donde residen las aplicaciones en las cuales trabajan, que pueden ser asociados a las plantillas que crea el usuario.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
Active	bit	1	1	0
Description	varchar	100	0	0
IDOS	smallint	2	5	0
IDOSTechnology	smallint	2	5	0
OSVersion	varchar	50	0	0
Priority	tinyint	1	3	0

Esquema: Reference

Nombre: OSTechnologies

Descripción: Tabla de referencia utilizada para almacenar las diferentes tecnologías de sistemas operativos (32, 64 bits, etc.). Esta información se asocian a los sistemas operativos disponibles para asociar a las plantillas que crea el usuario.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
---------	------	----------	-----------	--------

Active	bit	1	1	0
Description	varchar	50	0	0
IDOSTechnology	smallint	2	5	0

Esquema: Reference

Nombre: SWArchitectures

Descripción: Tabla de referencia que contiene las diferentes arquitecturas a nivel de software que pueden ser utilizados por las plantillas que crea el usuario.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
Active	bit	1	1	0
IDSWArchitecture	int	4	10	0
Priority	tinyint	1	3	0
SWArchitecture	varchar	100	0	0

Esquema: Reference

Nombre: Templates

Descripción: Tabla que consolida toda la información de las plantillas ingresadas por los usuarios.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
Active	bit	1	1	0
ClusterID	int	4	10	0
IDDBEngine	smallint	2	5	0
IDDevEnv	smallint	2	5	0
IDDevLanguage	smallint	2	5	0
IDOS	smallint	2	5	0
IDTemplate	int	4	10	0
IDVirtualizationOption	smallint	2	5	0
IDWebServer	smallint	2	5	0
LastModifiedDate	smalldatetime	4	16	0
TemplateName	varchar	150	0	0
UserIDSID	varchar	100	0	0
Version	int	4	10	0

Esquema: Reference

Nombre: Templates_DBArchitectures

Descripción: Tabla intermedia utilizada para manejar las relaciones n:n entre las plantillas y las arquitecturas de bases de datos.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
IDDBArchitecture	int	4	10	0

IDTemplate	int	4	10	0
IDTemplateDBArch	bigint	8	19	0
TemplateVersion	int	4	10	0

Esquema: Reference

Nombre: Templates_HWArchitectures

Descripción: Tabla intermedia utilizada para manejar las relaciones n:n entre las plantillas y las arquitecturas de hardware.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
IDHWArchitecture	int	4	10	0
IDTemplate	int	4	10	0
IDTemplateHWArch	bigint	8	19	0
TemplateVersion	int	4	10	0

Esquema: Reference

Nombre: Templates_SWArchitectures

Descripción: Tabla intermedia utilizada para manejar las relaciones n:n entre las plantillas y las arquitecturas de software.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
IDSWArchitecture	int	4	10	0
IDTemplate	int	4	10	0
IDTemplateSWArch	bigint	8	19	0
TemplateVersion	int	4	10	0

Esquema: Reference

Nombre: Templates_ThirdPartyComponents

Descripción: Tabla intermedia utilizada para manejar las relaciones n:n entre las plantillas y los componentes de terceros.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
IDTemplate	int	4	10	0
IDTemplatesThirdPty	bigint	8	19	0
IDThirdPtyComponent	int	4	10	0
TemplateVersion	int	4	10	0

Esquema: Reference

Nombre: ThirdPartyComponents

Descripción: Tabla de referencia que contiene los diferentes componentes de terceros que pueden ser utilizados por las plantillas que crea el usuario.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
Active	bit	1	1	0

ComponentName	varchar	150	0	0
IDThirdPtyComponent	int	4	10	0
Priority	tinyint	1	3	0
Version	varchar	50	0	0

Esquema: Reference

Nombre: VirtualizationOptions

Descripción: Tabla de referencia que contiene las diferentes opciones de virtualización que pueden ser utilizados por las plantillas que crea el usuario.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
Active	bit	1	1	0
IDVirtualizationOption	smallint	2	5	0
Priority	tinyint	1	3	0
VirtualizationOption	varchar	100	0	0

Esquema: Reference

Nombre: WebServers

Descripción: Tabla de referencia que contiene las diferentes opciones de servidores web que pueden ser utilizados por las plantillas que crea el usuario.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
Active	bit	1	1	0
IDWebServer	smallint	2	5	0
Priority	tinyint	1	3	0
Version	varchar	50	0	0
WebServer	varchar	100	0	0

Esquema: Staging

Nombre: tmpTemplateClusters

Descripción: Tabla de almacenamiento volátil, utilizada cuando se ejecuta la tarea automatizada de agrupamiento de plantillas. La tabla almacena los resultados de la consulta al modelo de minería de datos para cada una de las plantillas y sus versiones.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
ClusterID	int	4	10	0
IDTemplate	int	4	10	0
Version	int	4	10	0

Vistas

Esquema: Production

Nombre: vTemplateInfo

Descripción: vista que consolida toda la información de las plantillas. Une la tabla principal de plantillas con las tablas de referencia y presenta toda la información.

Columnas:

Columna	Tipo	Longitud	Precision	Escala
ClusterID	int	4	10	0
ComponentVersion	varchar	50	0	0
DatabaseEngine	varchar	100	0	0
DBArchitecture	varchar	100	0	0
DBVersion	varchar	100	0	0
DevEnvironment	varchar	150	0	0
DevEnvVersion	varchar	50	0	0
DevLanguage	varchar	100	0	0
DevLangVersion	varchar	50	0	0
HWArchitecture	varchar	100	0	0
IDDBArchitecture	int	4	10	0
IDHWArchitecture	int	4	10	0
IDSWArchitecture	int	4	10	0
IDTemplate	int	4	10	0
IDThirdPtyComponent	int	4	10	0
OS	varchar	100	0	0
OSTechnology	varchar	50	0	0
OSVersion	varchar	50	0	0
SWArchitecture	varchar	100	0	0
TemplateName	varchar	150	0	0
ThirdPtyComponent	varchar	150	0	0
Version	int	4	10	0
VirtualizationOption	varchar	100	0	0
WebServer	varchar	100	0	0
WebServerVersion	varchar	50	0	0

Procedimientos Almacenados

Esquema: Production

Nombre: AddComment

Descripción: Agrega un comentario a un hallazgo previamente ingresado por un usuario.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@CommentText	varchar	-1	0	0
@IDFinding	bigint	8	19	0
@SubmittedBy	varchar	100	0	0

Esquema: Production

Nombre: AddFinding

Descripción: Agrega un hallazgo al ítem investigado que se especifica en los parámetros.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@Details	varchar	-1	0	0
@IDResearchedItem	int	4	10	0
@User	varchar	100	0	0

Esquema: Production

Nombre: AddResearchedItem

Descripción: Agrega un ítem investigado por un usuario, asociado a una plantilla.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@ErrorMsg	varchar	150	0	0
@IDTemplate	int	4	10	0
@ResearchDescription	varchar	-1	0	0
@User	varchar	100	0	0

Esquema: Production

Nombre: GetAdvancedSearchResult

Descripción: Recibe del modelo de minería de datos el número de grupo al cual pertenecería la plantilla según las características seleccionadas por el usuario, y combina esto con los resultados de la búsqueda semántica por descripción o mensaje de error para habilitar la característica de "búsqueda avanzada" dentro de la herramienta.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@ClusterID	int	4	10	0
@Description	varchar	2000	0	0
@Message	varchar	2000	0	0

Esquema: Production

Nombre: GetBasicSearchResult

Descripción: Ejecuta la consulta semántica sobre las tablas con los ítems investigados para habilitar la función de "búsqueda básica" en la herramienta.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@Description	varchar	2000	0	0
@Message	varchar	2000	0	0

Esquema: Production

Nombre: GetFinding

Descripción: Consulta la base de datos en búsqueda de un hallazgo específico.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDFinding	bigint	8	19	0

Esquema: Production

Nombre: GetFindingComments

Descripción: Consulta la base de datos en búsqueda de los comentarios asociados a un hallazgo específico.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDFinding	bigint	8	19	0

Esquema: Production

Nombre: GetMyPosts

Descripción: Consulta la base de datos en búsqueda de los ítems agregados por un usuario específico.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@Submitter	varchar	100	0	0

Esquema: Production

Nombre: GetPostFindings

Descripción: Recupera los hallazgos relacionados con un ítem investigado específico.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDResearchedItem	int	4	10	0

Esquema: Production

Nombre: GetPostInfo

Descripción: Recupera los detalles de un ítem investigado específico.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
-----------	------	----------	-----------	--------

@IDResearchItem	int	4	10	0
-----------------	-----	---	----	---

Esquema: Production

Nombre: GetSimilarRIFromText

Descripción: Utiliza una consulta semántica para buscar ítems investigados con palabras claves similares al ítem especificado en los parámetros.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDResearchItem	int	4	10	0

Esquema: Production

Nombre: GetSpecificResearchedItem

Descripción: Consulta la base de datos en búsqueda de un ítem investigado específico.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDResearchItem	int	4	10	0

Esquema: Production

Nombre: GetTemplate_DBArch

Descripción: Recupera las Arquitecturas de Bases de Datos asociadas a una plantilla específica.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDTemplate	int	4	10	0

Esquema: Production

Nombre: GetTemplate_HWArch

Descripción: Recupera las Arquitecturas de Hardware asociadas a una plantilla específica.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDTemplate	int	4	10	0

Esquema: Production

Nombre: GetTemplate_SWArch

Descripción: Recupera las Arquitecturas de Software asociadas a una plantilla específica.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDTemplate	int	4	10	0

Esquema: Production

Nombre: GetTemplate_ThirdPty

Descripción: Recupera los componentes de terceros asociados a una plantilla específica.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDTemplate	int	4	10	0

@IDTemplate	int	4	10	0
-------------	-----	---	----	---

Esquema: Production

Nombre: GetTemplateDetails

Descripción: Consulta la base de datos y recupera la información de una plantilla específica.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDTemplate	int	4	10	0

Esquema: Production

Nombre: ReceiveFindingFeedback

Descripción: Ingresa la retroalimentación de un usuario acerca de un hallazgo publicado en la aplicación. Si el usuario ya emitió la retroalimentación sobre ese hallazgo, actualiza el registro. De lo contrario, lo agrega.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDFinding	bigint	8	19	0
@Positive	bit	1	1	0
@User	varchar	100	0	0

Esquema: Reference

Nombre: AddDBArchitecture

Descripción: Agrega una Arquitectura de Bases de Datos nueva al catálogo de referencia, para posteriormente ser utilizado como característica de una plantilla.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@DBArchitecture	varchar	100	0	0

Esquema: Reference

Nombre: AddDBEngine

Descripción: Agrega un Motor de Bases de Datos nuevo al catálogo de referencia, para posteriormente ser utilizado como característica de una plantilla.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@DBEngine	varchar	100	0	0
@Version	varchar	100	0	0

Esquema: Reference

Nombre: AddDevEnvironment

Descripción: Agrega un ambiente de desarrollo nuevo al catálogo de referencia, para posteriormente ser utilizado como característica de una plantilla.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
-----------	------	----------	-----------	--------

@Environment	varchar	150	0	0
@Version	varchar	50	0	0

Esquema: Reference

Nombre: AddDevLanguage

Descripción: Agrega un lenguaje de desarrollo nuevo al catálogo de referencia, para posteriormente ser utilizado como característica de una plantilla.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@Language	varchar	100	0	0
@Version	varchar	50	0	0

Esquema: Reference

Nombre: AddHWArchitecture

Descripción: Agrega una arquitectura de hardware nueva al catálogo de referencia, para posteriormente ser utilizado como característica de una plantilla.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@HWArchitecture	varchar	100	0	0

Esquema: Reference

Nombre: AddOperatingSystem

Descripción: Agrega un sistema operativo nuevo al catálogo de referencia, para posteriormente ser utilizado como característica de una plantilla.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@OS	varchar	100	0	0
@OSTechnology	tinyint	1	3	0
@Version	varchar	50	0	0

Esquema: Reference

Nombre: AddSWArchitecture

Descripción: Agrega una arquitectura de software nueva al catálogo de referencia, para posteriormente ser utilizado como característica de una plantilla.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@SWArchitecture	varchar	100	0	0

Esquema: Reference

Nombre: AddTemplate

Descripción: Agrega una plantilla nueva a la base de datos.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
-----------	------	----------	-----------	--------

@ClusterID	int	4	10	0
@DBEngine	smallint	2	5	0
@IDDevEnv	smallint	2	5	0
@IDDevLanguage	smallint	2	5	0
@IDOS	smallint	2	5	0
@IDVirtualizationOption	smallint	2	5	0
@IDWebserver	tinyint	1	3	0
@Name	varchar	150	0	0
@User	varchar	100	0	0

Esquema: Reference

Nombre: AddTemplate_DBArch

Descripción: Asocia una plantilla a una arquitectura de bases de datos.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDDBArchitecture	int	4	10	0
@IDTemplate	int	4	10	0

Esquema: Reference

Nombre: AddTemplate_HWArch

Descripción: Asocia una plantilla a una arquitectura de hardware.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDHWArchitecture	int	4	10	0
@IDTemplate	int	4	10	0

Esquema: Reference

Nombre: AddTemplate_SWArch

Descripción: Asocia una plantilla a una arquitectura de software.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDSWArchitecture	int	4	10	0
@IDTemplate	int	4	10	0

Esquema: Reference

Nombre: AddTemplate_ThirdPty

Descripción: Asocia una plantilla a un componente de terceros.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDComponent	int	4	10	0
@IDTemplate	int	4	10	0

Esquema: Reference

Nombre: AddThirdPtyComponent

Descripción: Agrega un componente de terceros nuevo al catálogo de referencia, para posteriormente ser utilizado como característica de una plantilla.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@Component	varchar	150	0	0
@Version	varchar	50	0	0

Esquema: Reference

Nombre: AddVirtualizationOption

Descripción: Agrega una opción de virtualización nueva al catálogo de referencia, para posteriormente ser utilizado como característica de una plantilla.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@VirtualizationOption	varchar	100	0	0

Esquema: Reference

Nombre: AddWebServer

Descripción: Agrega un servidor web nuevo al catálogo de referencia, para posteriormente ser utilizado como característica de una plantilla.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@Version	varchar	50	0	0
@WebServer	varchar	100	0	0

Esquema: Reference

Nombre: GetMyTemplates

Descripción: Consulta la base de datos y recupera la información de las plantillas activas creadas por el usuario especificado.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@UserID	varchar	100	0	0

Esquema: Reference

Nombre: RemoveComment

Descripción: Elimina el comentario asociado a un hallazgo que se especifica en el parámetro

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDComment	bigint	8	19	0

Esquema: Reference

Nombre: RemoveDBArchitecture

Descripción: Desactiva la arquitectura de bases de datos especificada. Esta no podrá ser utilizada por ninguna plantilla más pero las plantillas que ya la referencian no se afectan.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDDBArchitecture	smallint	2	5	0

Esquema: Reference

Nombre: RemoveDBEngine

Descripción: Desactiva el motor de bases de datos especificado. Este no podrá ser utilizado por ninguna plantilla más pero las plantillas que ya lo referencian no se afectan.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDDBEngine	smallint	2	5	0

Esquema: Reference

Nombre: RemoveDevEnvironment

Descripción: Desactiva el ambiente de desarrollo especificado. Este no podrá ser utilizado por ninguna plantilla más pero las plantillas que ya lo referencian no se afectan.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDDevEnvironment	smallint	2	5	0

Esquema: Reference

Nombre: RemoveDevLanguage

Descripción: Desactiva el ambiente de desarrollo especificado. Este no podrá ser utilizado por ninguna plantilla más pero las plantillas que ya lo referencian no se afectan.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDDevLanguage	smallint	2	5	0

Esquema: Reference

Nombre: RemoveFinding

Descripción: Elimina el hallazgo especificado.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDFinding	int	4	10	0

Esquema: Reference

Nombre: RemoveFindingFeedback

Descripción: Elimina la retroalimentación recolectada por un usuario específico sobre un hallazgo.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDFinding	int	4	10	0
@IDUser	varchar	100	0	0

Esquema: Reference

Nombre: RemoveHWArchitecture

Descripción: Desactiva la arquitectura de hardware especificada. Esta no podrá ser utilizada por ninguna plantilla más pero las plantillas que ya la referencian no se afectan.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDHWArchitecture	smallint	2	5	0

Esquema: Reference

Nombre: RemoveOS

Descripción: Desactiva el sistema operativo especificado. Este no podrá ser utilizado por ninguna plantilla más pero las plantillas que ya lo referencian no se afectan.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDOS	smallint	2	5	0

Esquema: Reference

Nombre: RemoveResearchedItem

Descripción: Elimina un ítem investigado así como toda la información asociada a él (comentarios, retroalimentación, etc.)

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDResearchedItem	int	4	10	0

Esquema: Reference

Nombre: RemoveSWArchitecture

Descripción: Desactiva la arquitectura de software especificada. Esta no podrá ser utilizada por ninguna plantilla más pero las plantillas que ya la referencian no se afectan.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDSWArchitecture	smallint	2	5	0

Esquema: Reference

Nombre: RemoveTemplate

Descripción: Desactiva una plantilla. Esta no podrá ser utilizada por el usuario nuevamente, pero toda su información se mantiene como información para entrenar el modelo de minería de datos.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDTemplate	int	4	10	0

Esquema: Reference

Nombre: RemoveThirdPtyComponent

Descripción: Desactiva el componente de terceros especificado. Este no podrá ser utilizado por ninguna plantilla más pero las plantillas que ya lo referencian no se afectan.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDComponent	smallint	2	5	0

Esquema: Reference

Nombre: RemoveVirtualizationOptions

Descripción: Desactiva la opción de virtualización especificada. Esta no podrá ser utilizada por ninguna plantilla más pero las plantillas que ya la referencian no se afectan.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDVirtualizationOptions	smallint	2	5	0

Esquema: Reference

Nombre: RemoveWebServer

Descripción: Desactiva el servidor web especificado. Este no podrá ser utilizado por ninguna plantilla más pero las plantillas que ya lo referencian no se afectan.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDWebServer	tinyint	1	3	0

Esquema: Reference

Nombre: UpdateTemplate

Descripción: Actualiza una plantilla. No actualiza el registro existente sino que crea un registro nuevo en la tabla, incrementando el valor de la columna "versión" a la vez que desactiva cualquier otro registro en la tabla para esa plantilla (excepto el recién ingresado).

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@DBEngine	smallint	2	5	0
@IDDevEnv	smallint	2	5	0
@IDDevLanguage	smallint	2	5	0
@IDOS	smallint	2	5	0
@IDTemplate	int	4	10	0
@IDVirtualizationOption	smallint	2	5	0
@IDWebserver	tinyint	1	3	0
@Name	varchar	150	0	0
@User	varchar	100	0	0

Funciones

Esquema: Production

Nombre: GetCommentUser

Descripción: Consulta la base de datos y retorna el usuario que ingresó en comentario especificado en el parámetro.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDComment	bigint	8	19	0

Esquema: Production

Nombre: GetFindingFeedback

Descripción: Consulta la base de datos y retorna las estadísticas de retroalimentación (totales positivos y negativos, porcentajes, etc.) del hallazgo especificado en el parámetro.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDFinding	bigint	8	19	0

Esquema: Production

Nombre: GetFindingUser

Descripción: Consulta la base de datos y retorna el usuario que ingresó el hallazgo especificado en el parámetro.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDFinding	bigint	8	19	0

Esquema: Production

Nombre: GetItemUser

Descripción: Consulta la base de datos y retorna el usuario que ingresó el ítem investigado que se especifica en el parámetro.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDResearchedItem	int	4	10	0

Esquema: Reference

Nombre: GetDBEngineInfo

Descripción: Consulta la base de datos y retorna la información del motor de bases de datos especificado en el parámetro.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDDDBEngine	int	4	10	0

Esquema: Reference

Nombre: GetDevEnvironmentInfo

Descripción: Consulta la base de datos y retorna la información del ambiente de desarrollo especificado en el parámetro.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDDevEnv	int	4	10	0

Esquema: Reference

Nombre: GetDevLanguageInfo

Descripción: Consulta la base de datos y retorna la información del lenguaje de programación especificado en el parámetro.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDDevLang	int	4	10	0

Esquema: Reference

Nombre: GetOSInfo

Descripción: Consulta la base de datos y retorna la información del sistema operativo especificado en el parámetro.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDOS	int	4	10	0

Esquema: Reference

Nombre: GetWebServerInfo

Descripción: Consulta la base de datos y retorna la información del servidor web especificado en el parámetro.

Parámetros:

Parametro	Tipo	Longitud	Precision	Escala
@IDWebServer	int	4	10	0

Bibliografía

APTE, Uday (et al). *Knowledge Management and Organizational Learning*. King, William R. (ed. lit.). Nueva York: Springer Science+Business Media, LLC. 2009. 396 p. ISBN: 978-1-4419-0007-4.

BARONI, Rodrigo; ARAÚJO, Marta. "Knowledge Management Software". En: SCHWARTZ, David G. *Encyclopedia of Knowledge Management*. Pensilvania: Idea Group Inc., 2006. p410 - 418. ISBN: 1-59140-574-2.

BECERRA-FERNANDEZ, Irma; SABHERWAL, Rajiv. *Knowledge Management Systems and Processes*. Gibson, Cyrus F (prol). Nueva York: M. E. Sharp, Inc., 2010. 352 p. ISBN: 978-0-7656-2351-5.

BOLAÑOS, Eduardo. Entrevista personal realizada el Lunes 26 de marzo del 2012, 12 pm.

CAVAZOS, Adriana. Entrevista personal realizada el Martes 20 de marzo del 2012, 3 pm.

CALVO, Ricardo. Entrevista personal realizada el Jueves 29 de marzo del 2012, 8:30 am.

CEPEDA-CARRIÓN, Gabriel. "Competitive Advantage of Knowledge Management". En: SCHWARTZ, David G. *Encyclopedia of Knowledge Management*. Pensilvania: Idea Group Inc., 2006. p34 - 43. ISBN: 1-59140-574-2.

FIGALLO, Cliff; RHINE, Nancy. *Building the Knowledge Management Network*. Nueva York: John Wiley & Sons, Inc., 2002. 348p. ISBN: 0-471-21549-X.

GOLDSMITH, Ronald E.; GOPALAKRISHNA, Kishore. "Knowledge Calibration". En: SCHWARTZ, David G. *Encyclopedia of Knowledge Management*. Pensilvania: Idea Group Inc., 2006. p311 - 316. ISBN: 1-59140-574-2.

JENNEX, Murray E. "Knowledge Management System Success Factors". En: SCHWARTZ, David G. *Encyclopedia of Knowledge Management*. Pensilvania: Idea Group Inc., 2006. p429 - 435. ISBN: 1-59140-574-2.

LAND, Frank (et al). "Theoretical and Practical Aspects of Knowledge Management". En: SCHWARTZ, David G. *Encyclopedia of Knowledge Management*. Pensilvania: Idea Group Inc., 2006. p410 - 418. ISBN: 1-59140-574-2.

LINDSEY, Keith L. "Knowledge Sharing Barriers". En: SCHWARTZ, David G. *Encyclopedia of Knowledge Management*. Pensilvania: Idea Group Inc., 2006. p499-506. ISBN:1-59140-574-2.

MAIER, Ronald; HÄDRICH, Thomas. "Knowledge Management Systems". En: SCHWARTZ, David G. *Encyclopedia of Knowledge Management*. Pensilvania: Idea Group Inc., 2006. p442 - 451. ISBN: 1-59140-574-2.

MORA, Wayler. Entrevista personal realizada el Miércoles 14 de marzo del 2012, 11 am.

WILSON, Rick L. (et al). "Knowledge Structure and Data Mining Techniques". En: SCHWARTZ, David G. *Encyclopedia of Knowledge Management*. Pensilvania: Idea Group Inc., 2006. p523 - 529. ISBN: 1-59140-574-2.

Rodríguez, Oldemar. Material didáctico del curso “PF-3845 Metodos y Algoritmos en Minería de Datos” de la Universidad de Costa Rica.