

Estadísticas distribuidas y gestión de calidad en el marco de la Iniciativa de Enlace de Bibliotecas

Abstract - La Iniciativa de Enlace de Bibliotecas (LibLink) del Consorcio Iberoamericano para la Educación en Ciencia y Tecnología (ISTEC) es un proyecto global que ha mostrado un crecimiento sostenido en los últimos años, acompañado de una permanente actividad por parte de sus miembros en busca del mejoramiento de los servicios, el incremento de la comunicación entre los participantes y la incorporación continua de nuevos miembros. En este sentido, los beneficios de contar con un sistema global de generación de estadísticas resultarán de gran utilidad tanto para corroborar la calidad del servicio -por ejemplo en lo relativo a la velocidad de atención de usuarios-, cuanto en la demostración del ahorro en cada entidad académica por el material obtenido gratuitamente de la red de cooperantes en lugar de por adquisición. Un beneficio adicional a mencionar es la posibilidad de obtener los títulos más solicitados en las distintas entidades y realizar compras consorciadas a menor costo.

El uso de una plataforma única como el software Celsius para la gestión de los pedidos bibliográficos de usuarios, con un fuerte hincapié en la generación de estadísticas locales de cada Institución, ofrece una oportunidad única para lograr una red distribuida de participantes de LibLink permitiendo, entre otros, automatizar procesos mediante la intercomunicación de instancias de Celsius, aumentar la comunicación entre todos los miembros y, principalmente, obtener datos propios de cada Institución participante para generar estadísticas globales que involucren a todos los miembros. En este artículo se detallan los mecanismos y tecnologías utilizadas para generar estadísticas distribuidas, y las posibilidades que esto ofrece. En la primera sección se muestra brevemente cómo ha evolucionado el software Celsius desde su nacimiento hasta la versión actual, Celsius Network, que ha incorporado estas nuevas funciones; en la segunda sección se detalla el funcionamiento de la red Celsius Network, los componentes que intervienen y los mecanismos y tecnologías que se utilizan para lograr una comunicación directa entre todas las instancias; a continuación se pone el foco sobre este nuevo tipo de estadísticas, desde la obtención remota hasta su visualización en línea, y se muestra cómo está funcionando esto en la actualidad. Finalmente, se estudian las posibilidades que esta red distribuida ofrece más

allá de las funciones y estadísticas implementadas hasta el momento.

I. INTRODUCCIÓN

LA obtención de métricas es un requisito fundamental para que las organizaciones puedan conocer su funcionamiento y propender a mejorar la calidad de sus procesos[1]. Los servicios deben ser evaluados constantemente, buscando alcanzar estándares de calidad aceptables, optimizando los recursos disponibles y brindados, y conociendo en detalle las preferencias de los usuarios alcanzando a fin de predecir su comportamiento, anticiparse a sus acciones y ofrecer prestaciones más eficientes.

La Iniciativa de Enlace de Bibliotecas [2] del Consorcio Iberoamericano de Educación en Ciencia y Tecnología [3] representa un caso particular en cuanto al estudio de su funcionamiento se refiere. A diferencia de otros proyectos, LibLink posee una serie de características que hacen de la obtención de métricas una tarea de considerable complejidad, a saber:

- Los participantes de LibLink incluyen Bibliotecas, Universidades e Institutos, distribuidos geográficamente a lo largo del continente Americano y España
- Las posibilidades de acceso a infraestructura y tecnologías modernas entre todos los participantes es muy dispar
- Si bien la Iniciativa sigue los lineamientos de un plan estratégico delineados por su dirección, los participantes son colaboradores voluntarios y dependientes de las condiciones propias de cada institución
- Las realidades económicas son muy diferentes
- Sumado a las distancias geográficas, la coordinación de tareas conjuntas implica esfuerzos adicionales para sincronizar a los grupos de trabajo y alcanzar resultados reales, como ser:

1 M.R. De Giusti; Investigador adjunto sin director de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CICPBA) y Director del Proyecto de Enlace de Bibliotecas de la UNLP; email: marisa.degiusti@sedici.unlp.edu.ar

2 G.L. Villarreal; Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y Proyecto de Enlace de Bibliotecas de la UNLP; email: gonetil@sedici.unlp.edu.ar

3 A. Sobrado; Proyecto de Enlace de Bibliotecas de la UNLP; asobrado@sedici.unlp.edu.ar

4 A. Lira; Proyecto de Enlace de Bibliotecas de la UNLP; alira@sedici.unlp.edu.ar

- unificar idiomas (inglés, portugués, castellano)
- coordinar horarios y reuniones remotas
- adoptar metodologías comunes de trabajo

Más allá de los problemas que se presentan, el uso de las Tecnologías de Información y Comunicación permiten aminorar las distancias y simplificar considerablemente algunas tareas; la obtención de datos sobre el funcionamiento de la Iniciativa LibLink en cada uno de los miembros es un claro ejemplo de ello, gracias a la utilización del software Celsius Network para reunir los datos y acumularlos en un único lugar permite lograr una visión global de la iniciativa y una visión particular de cada uno de sus participantes.

A lo largo de este trabajo se describirán brevemente las facilidades que ofrece el software Celsius Network locales a cada institución, qué mecanismos se utilizan para recolectar y agrupar datos distribuidos y cómo se accede a estos datos en línea.

II. EL SOFTWARE CELSIUS

Celsius nació en el año 2001 como un software para generación de estadísticas para la iniciativa LibLink. Si bien originalmente fue pensado para mostrar datos relativos al funcionamiento de la iniciativa, el desarrollo derivó rápidamente a la configuración de una base de datos confiable y a la incorporación de herramientas de administración de solicitudes de usuarios locales y remotos, cuyo estado cambia a través de eventos disparados por administradores o usuarios del sistema.

Durante las pruebas locales, el software Celsius mostró ser muy versátil, útil y práctico para realizar la gestión de solicitudes, con lo cual se liberó para todos los participantes de LibLink; la utilización por parte de numerosos administradores tanto locales como del exterior llevó a la necesidad de incorporar nuevas funcionalidades. La instalación de Celsius instalada dentro de la red de cada miembro permitió la generación de estadísticas locales para cada institución simplificando una ardua tarea para los administradores y beneficiando a los usuarios con las facilidades que el software ofrecía. Con el correr de los años, la cantidad de participantes de LibLink creció de manera considerable, así como también la cantidad de *instancias* o copias de Celsius instaladas; mientras tanto, nuevas versiones de esta herramienta eran liberadas periódicamente, intentando satisfacer así a la cada vez mayor masa de usuarios.

El crecimiento de la iniciativa LibLink manifestado en la incorporación de nuevos miembros a su red de bibliotecas, hacía cada vez más complejo estudiar su funcionamiento y analizar puntos débiles y potenciales mejoras. Si bien

Celsius aportaba un módulo completo de estadísticas, la cantidad de instalaciones distribuidas por América y España había crecido demasiado, haciendo que la obtención de datos globales demandara una inversión de tiempo y esfuerzo excesivo. Para facilitar esta tarea – entre otras – se creó Celsius Network[4], la última versión de Celsius en la cual las instancias distribuidas se conectan entre sí y comparten información unas con otras.

III. CELSIUS NETWORK: TRAZANDO PUENTES ENTRE LOS MIEMBROS DE LIBLINK

Como ya se mencionó, Celsius Network (Celsius NT, versión 2.0) es la última versión de esta herramienta de gestión de solicitudes bibliográficas y cálculo de estadísticas. Esta rama del software – cuyo última versión liberada es la 2.0.3 – representa un cambio considerable en relación a las versiones anteriores de Celsius; entre sus principales características, se destacan:

- reingeniería de buena parte de sus módulos e implementación de mejoras solicitadas por los administradores y los usuarios
- automatización de procesos de administración remotos
- generación de estadísticas distribuidas

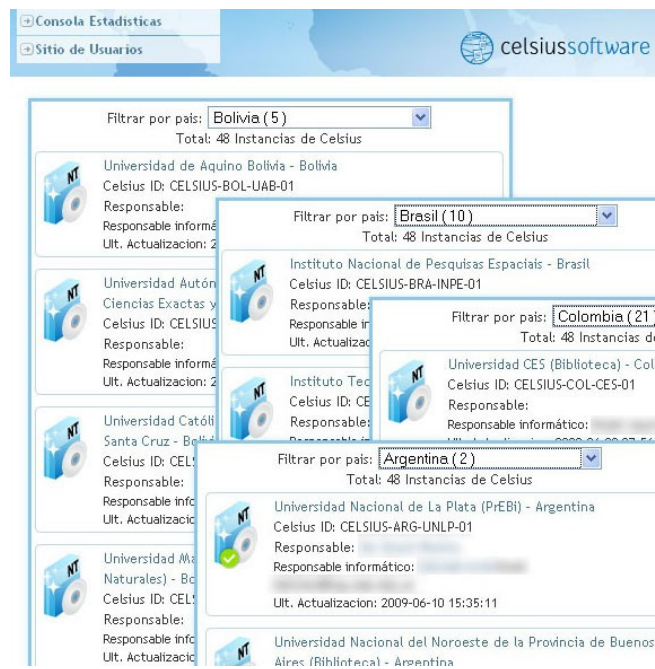


Fig. 1: Listado de Celsius instalados

Desde el lanzamiento de esta versión, el número de instituciones que han instalado y utilizan a diario esta herramienta ha crecido de algo menos de 30 a 48 instancias. El sitio web oficial del software Celsius[5] muestra un listado completo con las instituciones que poseen Celsius funcionando, junto con algunos datos importantes, entre los que se incluyen la versión de Celsius, el identificador global

único asignado a la institución, datos del responsable y datos del contacto informático a cargo. Si bien buena parte de esta información solo está disponible a usuarios registrados – a excepción de la versión y la url –, todas las instituciones con Celsius poseen al menos un usuario dentro de este sitio, con lo cual pueden obtener información muy valiosa para mantenerse en contacto con otros administradores Celsius.

La principal diferencia entre Celsius Network y las versiones anteriores de Celsius es su capacidad de conectarse directamente con otros Celsius NT ya sea para compartir información o para disparar procesos remotamente. Esto es posible principalmente gracias al uso de los servicios web mediante el protocolo HTTP [6] [7]. Para ello, cada instancia debe ahora saber que existen otras instancias, y debe disponer de información suficiente sobre de cada una de ellas – como ser dirección web, versión instalada y datos de acceso seguro – para localizarla y enviar o recibir información de manera segura y lo más transparente posible. Sumado a esto, cada instancia posee ahora un identificador global, que le permite generar solicitudes de usuario con un identificador único entre instancias, y que le permite a cualquier instancia conocer en cualquier momento el origen de un pedido a partir de su identificador. Finalmente, cada instancia debe conocer la estructura organizacional jerárquica en la cual se encuentra cada una de las otras instancias Celsius: País – Institución – Dependencia – Unidad – Instancia .

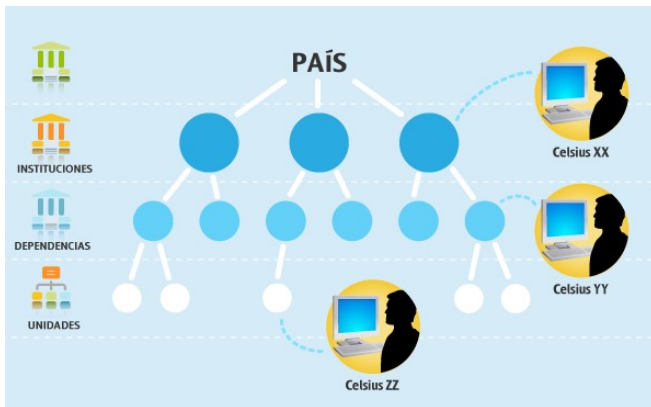


Fig. 2: Estructura jerárquica PIDUI

Compartir toda esta información resulta muy útil y productivo, pero a la vez puede tornarse muy complejo. Existe una serie de requerimientos mínimos provenientes de versiones previas, y otros que surgen a partir de la capacidad de las instancias de Celsius NT de conectarse unas con otras:

- Los cambios en una instancia particular deberían replicarse en todas las instancias
- La gestión de usuarios y claves de ingreso de cada instancia hacia otra instancia debe proveer un acceso

seguro, permitiendo un grado de flexibilidad elevado para poder incorporar nuevas instancias, modificar datos de ingreso o quitar instancias que ya no pertenezcan a la red

- La incorporación constante de nodos al árbol PIDUI debería diseminarse casi inmediatamente a todas las instancias
- En todo momento debe conocerse el estado de cada instancia y debe poder consultarse de manera rápida y sencilla (fecha de última actualización y resultado, su estado es activo o se ha caído, etcétera)
- La generación de estadísticas distribuidas debe realizarse en un espacio único en el que cada instancia pueda confiar al enviar sus estadísticas locales todas las cuales serán luego mostradas de un modo apropiado y sin sobrecargar a otras instancias con el procesamiento de datos y estadísticas que no sean propios
- Los administradores de redes y administradores Celsius deben contar con un espacio donde siempre puedan obtener las últimas versiones de este software, información para contactarse con otros usuarios, documentación y contacto con el equipo de desarrollo

Todos estos requerimientos fueron satisfechos mediante la incorporación de un nodo particular a la red Celsius NT: el directorio Celsius.

IV. El Directorio Celsius

El Directorio Celsius es una componente esencial dentro de la red Celsius Network, ya que posee la capacidad de centralizar, recolectar y diseminar toda la información que cada instancia de Celsius necesita para trabajar en conjunto con otras instancias. En este nuevo software también se han centralizado todas las funciones del Sitio web oficial del software Celsius, simplificando de este modo las tareas de actualización permanente y acceso a información por parte de administradores de redes.

Ahora, cada nuevo nodo Celsius NT se agrega a la red de instancias de manera muy rápida mediante su incorporación al directorio; si la incorporación del nuevo miembro requirió a su vez la actualización del árbol PIDUI, la administración del directorio Celsius permite hacerlo de manera muy sencilla, provocando una actualización de las instancias Celsius dirigida por el directorio, lo cual asegura que la información de cada instancia está siempre al día; asimismo, el directorio Celsius se encarga de administrar y difundir datos de acceso entre los Celsius NT, controlando que cada instancia de Celsius acceda a otras instancias con datos únicos (no repetibles) para cada par de instancias Celsius.

Con la incorporación del directorio Celsius, la centralización de estadísticas propias de cada instancia se realiza actualmente de un modo simple y rápido. Este directorio posee una sección especial, desde la cual se pueden acceder a las estadísticas de toda la red a través de sucesivos clics que indiquen el tipo de estadística, el período a analizar, la forma de generar los resultados y el tipo de gráfico a mostrar elegidos por el administrador.

Si bien para el usuario final resulta ser muy simple acceder a estadísticas centralizadas, el proceso completo es algo más complejo y requiere la intervención de varias componentes que permiten una recuperación rápida y segura de los datos, para luego mostrarlos apropiadamente:

1. Dada la cantidad de instancias Celsius existentes actualmente, y la clara tendencia hacia un crecimiento constante, es necesario contar con un mecanismo que permita al directorio solicitar a cada instancia datos propios de manera totalmente independiente del resto de las instancias; esto significa que las demoras en una instancia no deberían afectar a otras, y que el usuario final debería contar con los datos que ha requerido en un tiempo aceptablemente bueno. Es claro en este punto que la conexión entre el directorio y cada instancia de Celsius no puede realizarse en serie, sino que debe paralelizarse tanto como sea posible: desde el directorio se envía una petición simultánea a cada instancia activa, y se recuperan los datos enviados por cada una de manera concurrente. Una vez que todos los participantes han enviado sus datos (o han alcanzado un tiempo de espera máximo), el directorio podrá disponer de ellos de acuerdo a la selección del usuario. Pero existe un problema aquí: el directorio Celsius, así como también el software Celsius, fue desarrollado en el lenguaje de programación PHP 5. Este lenguaje de programación no fue pensado para realizar tareas en paralelo, con lo cual se hizo necesario buscar un método para paralelizar estas tareas, siempre atendiendo a no recargar con procesamiento adicional al cliente web. Para sortear este inconveniente, se ha utilizado la librería PHP-Java-Bridge[8]; esta herramienta permite ejecutar código Java dentro del lenguaje PHP, y se integra muy fácilmente a cualquier entorno de desarrollo preexistente. Este puente entre PHP y Java ha permitido al equipo de desarrollo generar una pequeña aplicación capaz de iniciar conexiones en paralelo mediante hilos o threads [9] desde el directorio hacia cada una de las instancias, y recibir datos mediante esta conexión dedicada sin afectar a otras instancias. El esquema de funcionamiento es el siguiente:

- Desde PHP, se recupera la lista de instancias junto con la

información de seguridad y localización necesaria para conectarse con cada una de ellas

- Se invoca al cliente multi-hilo desarrollado en Java, al cual se le envían todos estos datos previamente recolectados
- El cliente Java genera una conexión dedicada hacia cada instancia, mediante un thread único que se encarga de autenticarse ante cada instancia de Celsius para luego enviarle a la misma la petición de datos estadísticos, junto con todos los parámetros apropiados (consulta a realizar, período, formato de datos del resultado, ente otros).
- La instancia obtendrá los datos y realizará los cálculos apropiados según se le ha requerido, y los enviará nuevamente al cliente java
- El cliente java recibirá en paralelo los datos de todas las instancias, y cerrará las conexiones a medida que las instancias van respondiendo. En caso de tiempo de espera agotado, el cliente java simplemente ignorará a la instancia con lo cual la misma no aparecerá entre las estadísticas
- Una vez que todas las conexiones han finalizado, el cliente java enviará los datos recibidos nuevamente hacia la aplicación PHP, finalizando así su tarea
- Cuando la aplicación PHP ha obtenido la recepción completa, genera la página resultante de acuerdo a los parámetros que ha solicitado el usuario, siguiendo así el flujo normal de cualquier aplicación web.

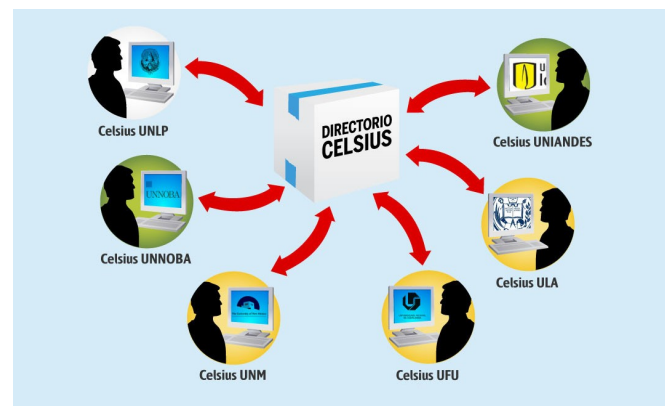


Fig. 3: Directorio Celsius e instancias Celsius Network

2. Desde el lanzamiento de Celsius Network (2.0, primera versión de Celsius Network), han sido liberadas hasta la fecha otras 3 nuevas versiones con nuevas incorporaciones, mejoras y soluciones a problemas detectados por los usuarios: Celsius NT 2.0.1, 2.0.2 y 2.0.3, y esto tiende a continuar. Con cada nueva versión, la estructura de la base de datos puede cambiar de manera considerable, para incorporar las nuevas funciones propias de la versión; esto hace que la solicitud de datos hacia cada instancia por parte del directorio deba realizarse teniendo en cuenta cada versión de Celsius NT, permitiendo así que puedan

participar de estas estadísticas centralizadas todas las instancias de Celsius Network sin importar la versión que poseen instalada hasta la fecha. Para lograr esto, el directorio Celsius envía peticiones específicas para cada versión de Celsius Network; dado que este directorio conoce de antemano qué versión se ha instalado en cada instancia, puede enviarle a las mismas una solicitud de estadísticas respetando su esquema propio de bases de datos actual y sin requerirle procesamiento adicional para normalizar los resultados de acuerdo a la estructura propia de una versión determinada. Todo este procesamiento queda a cargo del directorio Celsius, lo cual asegura la participación en las estadísticas tanto de nuevas versiones de Celsius Network como de versiones anteriores.

VI. ESTADÍSTICAS EXISTENTES

Hasta este momento se han generado algunas estadísticas que ya permiten obtener datos muy valiosos acerca del funcionamiento de cada instancia, de los requerimientos de los usuarios y del estado general de la iniciativa a nivel global:

- **Cantidad de Pedidos:** Obtiene la cantidad total de pedidos de todos los Celsius que participan en la red Celsius Network. Muestra los resultados por año y por mes; los datos son automáticamente agrupados por cada país de origen. Esto permite conocer la cantidad de solicitudes que son realizadas y/o satisfechas en un período determinado.
- **Tardanza promedio de pedidos:** Calcula la cantidad de días desde que los pedidos son realizados hasta que están disponibles. Esta estadística discrimina los resultados por instancia de Celsius NT, lo cual permite conocer qué Instituciones responden en tiempos mejores que el promedio, similares o por encima del mismo. Además de la tardanza promedio global, se calcula una tardanza promedio de cada institución, de modo de poder comparar los valores mensuales y/o anuales de las mismas con su propio promedio (y así conocer tendencias locales mensuales) o incluso la media global con cada promedio local.
- **Cantidad de usuarios:** Obtiene la cantidad de usuarios que se han incorporado mes a mes en cada instancia Celsius. Los resultados se muestran nuevamente discriminados por instancia, lo cual permite conocer cuáles son aquellas Instituciones que más han crecido en cantidad de usuarios en cada período. Para cada instancia, también se obtiene un promedio anual en base a los resultados mes a mes.
- **Principales títulos de cada Instancia.** Con esta

estadística se pueden conocer cuales son las revistas más solicitadas en cada una de las instancias participantes, ofreciendo información muy valiosa a la hora de realizar compras de materiales o suscripciones a revistas y/o catálogos.



Fig. 4: Directorio Celsius - estadísticas centralizadas

VI. CONCLUSIÓN Y TRABAJOS FUTUROS

Se han expuesto brevemente las facilidades que ofrece Celsius Network a la hora de recopilar estadísticas globales de utilidad en redes que comparten recursos bibliográficos. Es importante resaltar que este desarrollo, más allá de la innovación tecnológica que representa, sirve a un fin claro que ha sido paradigma de ISTECS: la gestión colaborativa de la información y la compartición de conocimientos en pos de un desarrollo equitativo y sostenible de Latinoamérica, es en este sentido que el esfuerzo dedicado al software Celsius, adquiere un mayor sentido y relevancia.

La propia naturaleza de la red que utiliza Celsius Network, su constante crecimiento, su perfil muy variable y el crecimiento sostenido de las áreas y los sistemas de información imponen un constante avance en los desarrollos informáticos para acompañar las nuevas necesidades de las bibliotecas y los requerimientos crecientes de usuarios cada

día más expertos, esto genera un compromiso en el desarrollo para ir acompañando los cambios.

Una función deseable del directorio Celsius es la de solicitar estadísticas de un conjunto de instancias, en vez de todas las instancias. Así, el usuario deberá marcar todas las instancias a las que desea consultar, y solo solo aquellas que han sido seleccionadas participarán en la estadística distribuída. Esto permite, por un lado, obtener estadísticas puntuales de una sola instancia; si bien esto se puede hacer ingresando al Celsius de esa instancia, se evitaría así ingresar a cada Celsius a consultar, uno por uno. Por el otro lado, esta funcionalidad permitirá obtener estadísticas regionales: podrían seleccionarse todas las instancias de un país, o de una región determinada, accediendo a datos más específicos y manejables; esto resulta especialmente conveniente cuando se realizan convenios o trabajos conjuntos entre instituciones, buscando obtener información de los participantes para tomar mejores decisiones conjuntas.

La generación de estadísticas se realiza actualmente en vivo, solicitando en el momento los datos a cada instancia y mostrando resultados de aquellas que han respondido dentro del tiempo estipulado. Podría suceder que por algún problema del momento (congestión en la red, caída temporal del servidor), alguna instancia no envíe sus resultados; para estos casos, el directorio Celsius podría contar con datos en una caché reciente[10], y mostrar esos datos. Si la caché se renueva en períodos suficientemente cortos, las estadísticas podrían aún mostrar datos reales mientras se maximiza la cantidad de instancias participantes.

Mantener los datos en caché puede incluso favorecer a disminuir la carga en las instancias al generar las estadísticas. En vez de enviar una solicitud directamente hacia la instancia cada vez que se le solicita, el directorio Celsius podría tener los datos locales previamente recolectados y mostrarlos sin tener que conectarse con ninguna instancia. Esto requerirá desde luego un mecanismo

de actualización periódica de estos datos, y presenta un desafío de diseño en cuanto a la forma de mantener estos datos, dando soporte a la heterogeneidad de las estadísticas existentes actualmente y permitiendo la incorporación de nuevas estadísticas al sistema.

VIII. REFERENCIAS

- [1] Charles R. McClure, R. David Lankes, Melissa Gross, Beverly Choltco-Devlin. *Statistics, Measures, and Quality Standards for Assessing Digital Reference Library Services: Guidelines and Procedures*, 2002
- [2] ISTECS Iniciativa de Enlace de Biblioteca (Library Linkage Initiative), 2009, agosto, <http://liblink.istec.org/>
- [3] Consorcio Iberoamericano para la Educación en Ciencia y Tecnología, 2009, agosto, <http://www.istec.org>
- [4] De Giusti, Marisa Raquel; Lira, Ariel Jorge; Inafuku Fernando Gabriel; Sobrado Ariel, Celsius Network, 2007
- [5] Sitio web oficial del software Celsius, 2009, agosto, <http://celsius.prebi.unlp.edu.ar>
- [6] De Giusti, Marisa Raquel; Vila, Maria Marta; Sobrado, Ariel; Villarreal, Gonzalo Lujan; Lira, Ariel Jorge. *Interconnection And Sharing Through Webservices*, 2007
- [7] Nammour, F. Mansour, N. *Comparative evaluation of object request broker technologies*, 2003
- [8] PHP/Java bridge, 2009
<http://php-java-bridge.sourceforge.net/pjb/>
- [9] Kwak, H. Lee, B. Hurson, A.R. Suk-Han Yoon Woo-Jong Hahn. *Effects of multithreading on cache performance*, 1999
- [10] Weikle, D.A.B. McKee, S.A. Wulf, W.A. *Caches as filters: a new approach to cache analysis.*, 1998