2DO. WORKSHOP NACIONAL_
Tecnologías e infraestructuras de la información para Bibliotecas y Centros de Documentación e Información
28_SEPTIEMBRE 2016
LANÚS | BUENOS AIRES

Aplicación para la obtención de indicadores de uso en repositorios institucionales

Expositor: Lic.Mariana Pichinini (FaHCE-UNLP)

mariana@fahce.unlp.edu.ar

#bibliotecasTIC
@comunidadsiu @CINoficial @caicyt

Auspician y colaboran con la organización del evento

Trabajo del Grupo Métricas del proyecto Investigación y Desarrollo en Repositorios Institucionales: aplicaciones y experiencias en universidades de la región bonaerense (PICTO-2010-0149 – 2012/2013).

Integrantes: Archuby, Gustavo; Caprile, Lorena; González, Claudia; Jorquera, Israel; Merlino, Cristian; Pichinini, Mariana; Romero, Roxana

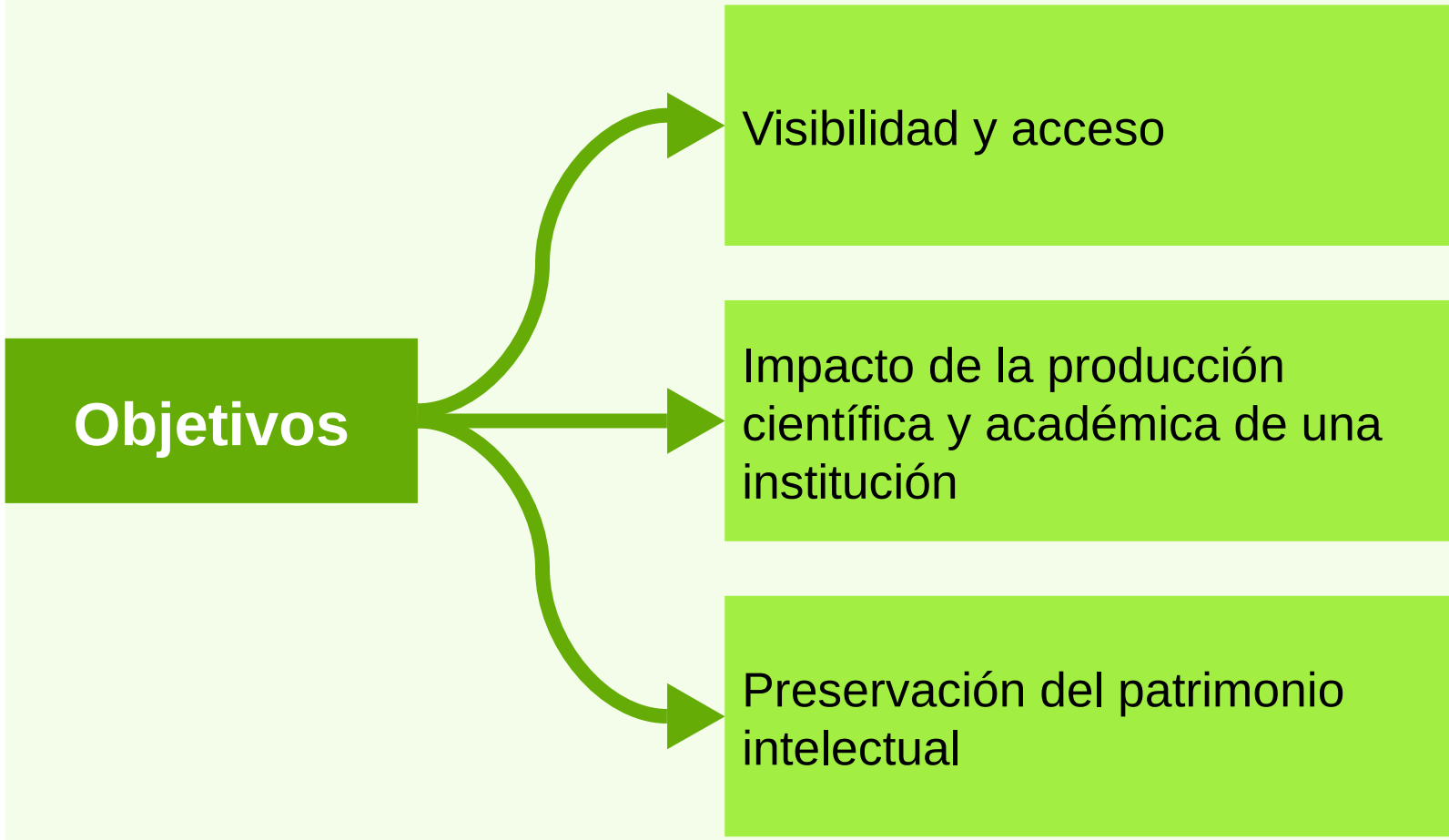
Resultados presentados en

Archuby, Gustavo; González, Claudia; Jorquera Vidal, Israel; Merlino, Cristian; Pichinini, Mariana. (2013). Medición de uso en repositorios digitales: Hacia la construcción de un marco de referencia argentino. III Jornadas de Intercambio y Reflexión acerca de la Investigación en Bibliotecología, 28 y 29 de noviembre de 2013, La Plata, Argentina. Disponible en: http://www.memoria.fahce.unlp.edu.ar/trab_eventos/ev.3363/ev.3363.pdf

Obtención de indicadores de uso en RI

RI

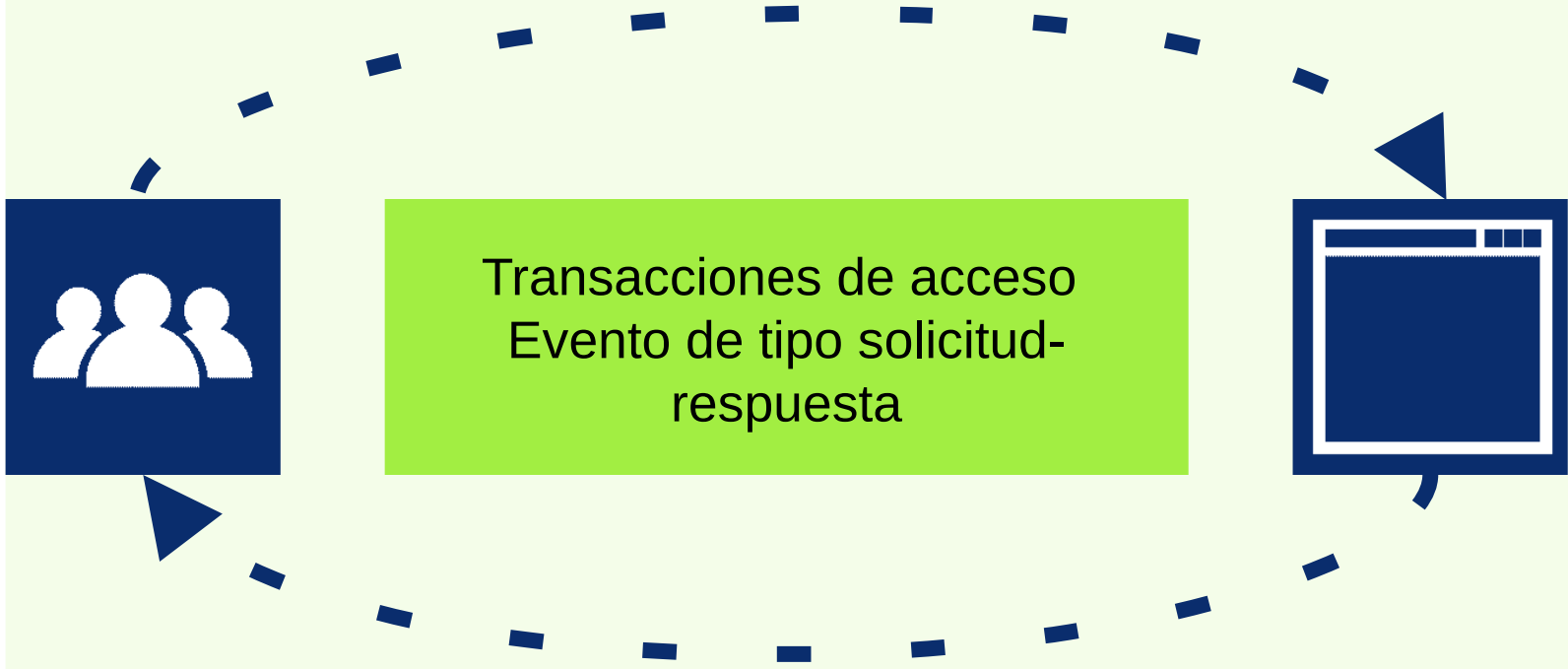
Repositorios Institucionales



Obtención de indicadores de uso en RI

Repositorios Institucionales

Necesidad de evidencia objetiva que refleje el uso de los contenidos digitales dispuestos en abierto.



Obtención de indicadores de uso en RI

Estadísticas de uso consistentes y confiables

Con múltiples repositorios es necesario implementar una práctica de trabajo consistente y homogénea con indicadores uniformes.

Obtención de indicadores de uso en RI

Objetivos

Establecer una práctica estandarizada de recolección y procesamiento de datos de uso de objetos digitales almacenados en repositorios institucionales

Definir un conjunto de indicadores que reflejen el uso de dichos objetos o sus representaciones

Obtener una aplicación que permita lograr los objetivos anteriores

Obtención de indicadores de uso en RI

Relevamiento de iniciativas internacionales y estado del arte

COUNTER (Counting Online Usage of Networked Electronic Resources)
EMIS (E-Metrics Instructional System).
KE (Knowledge Exchange).Guidelines for the Exchange of Usage Statistics.
MESUR (MEtrics from Scholarly Usage of Resources).
OA-Statistik.
PIRUS2 (Publisher and Institutional Repository Usage Statistics).
NISO Z39.7-201X, Information Services and Use: Metrics & statistics for libraries and information providers - Data Dictionary.
ISO/TR 20983:2003, Information and documentation - Performance indicators for electronic library services.

SURE 2 (Statistics on the Use of Repositories)

Obtención de indicadores de uso en RI

Relevamiento de formas posibles de obtención de datos

1) el uso de registros de archivos de transacciones, conocidos como **archivos de logs**.

2) las balizas web, llamada así por traducción de web beacon o también llamada web bug, técnica que se basa principalmente en la explotación de las **cookies de los navegadores**.

Obtención de indicadores de uso en RI

Relevamiento de formas posibles de obtención de datos

3) el uso de **JavaScript** incrustado en las páginas mediante etiquetas, la técnica preferida por las empresas que ofrecen servicios de Analítica Web.

4) lo que se conoce como packet sniffing, que implica la interposición entre el usuario y el servidor web de una **capa especial**, que puede ser de software o de hardware, que se encarga de recolectar la información.

Obtención de indicadores de uso en RI

Archivo de logs

Obtención de indicadores de uso en RI

Uso de archivos de registro de transacciones (logs)

- Registro de la totalidad de transacciones que se llevan a cabo entre los clientes y el servidor web
- Registra datos como fecha, dirección IP, URL que se solicitó, versión del protocolo, entre otros
- Formato estandarizado (Common Log Format CLF). Combined incluye datos del Referente.
- Independiente de las elecciones de seguridad y privacidad que establecen los usuarios al navegar

Obtención de indicadores de uso en RI

Requerimientos estándares de filtrado

- Código de estado de Hypertext Transfer Protocol 200 (OK) o 304 (No modificado)
- Tipo de petición GET
- El agente no sea un robot.
- URL corresponda a una ubicación del repositorio donde se almacenan los objetos digitales (ruta de directorio).
- En caso de haber dos accesos al mismo objeto desde una misma IP en menos de 30 segundos, solo se contará como 1 acceso.
- IPs inválidas (red local)

Obtención de indicadores de uso en RI

Indicadores - Niveles de medición

Primer nivel

Segundo nivel

Pueden obtenerse utilizando como fuente de datos los registros de transacciones, y por lo tanto, accesibles en cualquier institución que posea un repositorio disponible en la web

Obtención de indicadores de uso en RI

Indicadores - Niveles de medición

Primer nivel

Segundo nivel

Requieren además para su cómputo la obtención de información de otras fuentes, y por lo tanto, conllevan un nivel de complejidad mayor. El elemento que se usa en todos los casos para localizar esta información en otras fuentes es el URL del objeto digital (presente en la petición realizada al servidor web y registrada como transacción), que deberá servir como llave para su identificación.

Por ej., una base de datos con salida estándar o el protocolo OAI-PMH.

Obtención de indicadores de uso en RI

Indicadores - Niveles de medición

Primer nivel

Segundo nivel

Por estar definidos en función de los datos que se pueden obtener del registro de transacciones del servidor web y del protocolo OAI-PMH, estos indicadores son independientes del software utilizado para desarrollar cada repositorio, lo que garantiza su aplicabilidad en cualquier institución.

Obtención de indicadores de uso en RI

Indicadores - Niveles de medición

Primer nivel

Segundo nivel

Ambos grupos de indicadores deben calcularse en un rango determinado de fechas.

Obtención de indicadores de uso en RI

Indicadores de primer nivel

1. Total de descargas (TD)
2. Total de descargas directas (TDD)
3. Total de descargas locales (TDL)
4. Total de descargas desde buscadores (TDB)
5. Total de descargas desde otros servicios (TDO)
6. Total de descargas y visualizaciones por países (TDP)
7. Porcentaje de descargas por países (PDP)
8. Porcentajes de descargas de un país definido (PDPD)
9. Porcentajes de descargas de países no definidos (PDPND)
10. Porcentajes de descargas excepto de un país definido (PDEP)

Obtención de indicadores de uso en RI

Indicadores de primer nivel

11. Total de visualizaciones de registros bibliográficos (TVR)
12. Total de visualizaciones de registros bibliográficos por países (TVRP)
13. Porcentajes de visualizaciones de registros bibliográficos por países (PVRP)
14. Porcentajes de visualizaciones de registros bibliográficos de un país definido (PVRPD)
15. Porcentajes de visualizaciones de registros bibliográficos de países no definidos (PVRPND)
16. Porcentajes de visualizaciones de registros bibliográficos excepto de un país definido (PCREP)

Obtención de indicadores de uso en RI

Indicadores de segundo nivel

1. Número objetos digitales no descargados (NODND)
2. Total de descargas por tipo (TDT)
3. Total de descargas por tema principal (TDTP)
4. Total de descargas por idioma (TDI)
5. Total de descargas por fecha de publicación (TDFP)
6. Total de descargas por título de revista (TDTR)
7. Total de descargas por autor (TDA)
8. Total de descargas por versión del documento (TDV)
9. Total de descargas por filiación institucional (TDFI)

Obtención de indicadores de uso en RI

Desarrollo de la aplicación - Objetivo general

- Considerar los distintos niveles de desarrollo de los repositorios e intentar asegurar que todos los servicios puedan relevar el uso de los objetos digitales

- Utilizar tecnología estándar que permita escalar la aplicación con programación modular en respuesta a necesidades particulares

Obtención de indicadores de uso en RI

Desarrollo de la aplicación - Características

- Versión limitada a los indicadores de primer nivel que no requieren fuente externa de datos
- Adaptable a distintos entornos de trabajo
- Generación de indicadores en forma consistente y homogénea.
- Filtrado de los archivos de logs
- Base de datos con accesos válidos
- Cálculo de indicadores mediante consultas preseteadas

Obtención de indicadores de uso en RI

Desarrollo de la aplicación

- Entorno de programación Python 2.7.3
- Gestor de base de datos MySQL versión 14 o superior
- La aplicación se distribuye como un archivo comprimido denominado **indicadores.tar.gz**

Obtención de indicadores de uso en RI

Instalación y uso

- Descomprimir en una carpeta

- Inicializar la base MySQL

- Configurar los parámetros de procesamiento

- Importar los archivos de logs

- Realizar las consultas

Obtención de indicadores de uso en RI

Archivos de configuración

Tablas GeoIPLite para la geolocalización de IPs a nivel país

Listado de robots

(se utiliza la lista de COUNTER versión julio 2016)

Listado de search engines

(se utiliza la lista que provee PIWIK versión 2016)

Obtención de indicadores de uso en RI

Archivos de configuración

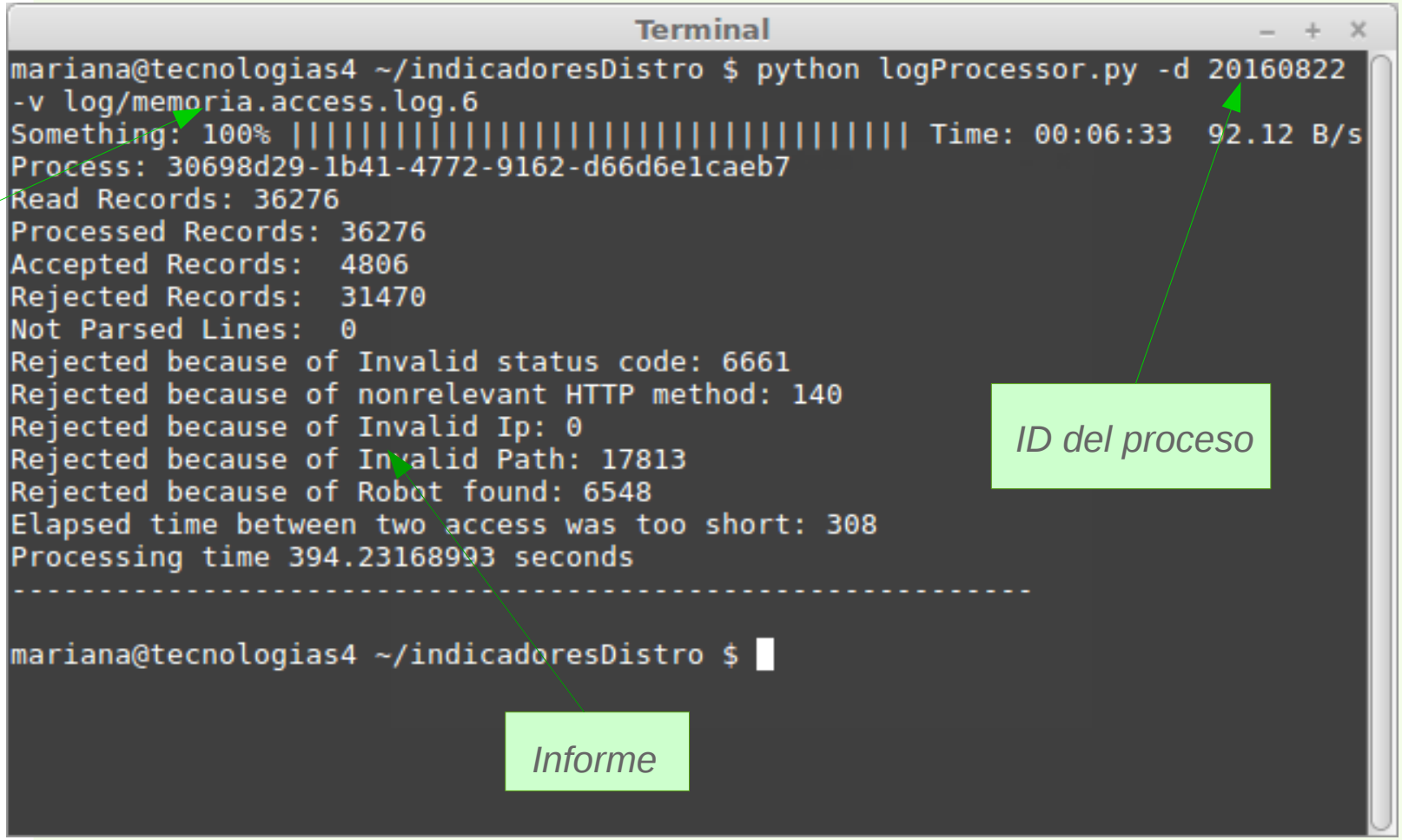
Expresiones regulares para la definición de paths válidos

Se incluyen ejemplos de paths válidos para repositorios desarrollados en Greenstone, Eprints y Dspace

Expresiones regulares para identificar diferencias entre descargas y visualizaciones de registros

Obtención de indicadores de uso en RI

Componentes: El procesador de logs



Obtención de indicadores de uso en RI

Componentes: La consulta de indicadores

Conexión con la base SQL

Consultas preseteadas

Resultado

Terminal

```
mariana@tecnologias4 ~/indicadoresDistro $ python
Python 2.7.12 (default, Jul  1 2016, 15:12:24)
[GCC 5.4.0 20160609] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> import indicadores
>>> q = indicadores.MySQLQueryProcessor("localhost", "root", "daniel2406", "indi
cadores")
>>> q.count_total_downloads("20160101", "20161231")
11828L
>>> q.count_downloads_from_search_engine("20160101", "20161231")
6998L
>>> q.percentage_of_downloads_from_country("20160101", "20161231", "AR")
26.32
>>> q.percentage_of_downloads_grouping_by_country("20160101", "20161231")
[('AR', 26.32), ('MX', 16.45), ('CO', 13.04), ('ES', 9.58), ('VE', 5.77), ('DO',
 5.24), ('US', 3.5), ('PE', 2.76), ('EC', 1.66), ('BO', 1.52), ('GT', 1.45), ('U
Y', 1.1), ('CL', 1.01), ('BR', 0.93), ('CR', 0.85), ('CN', 0.81), ('HN', 0.8), (
'SV', 0.74), ('GB', 0.66), ('DE', 0.64), ('IT', 0.5), ('AP', 0.48), ('FR', 0.48)
, ('NI', 0.39), ('PY', 0.35), ('EU', 0.3), ('PA', 0.25), ('PR', 0.19), ('IR', 0.
17), ('JP', 0.17), ('CU', 0.14), ('NL', 0.12), ('PH', 0.12), ('PT', 0.12), ('RU'
, 0.12), ('NG', 0.1), ('KH', 0.1), ('IN', 0.1), ('CA', 0.08), ('CH', 0.08), ('BE
', 0.08), ('PL', 0.08), ('IL', 0.08), ('DZ', 0.06), ('BG', 0.05), ('RO', 0.05),
('IE', 0.04), ('AD', 0.03), ('UA', 0.03), ('AT', 0.03), ('HR', 0.03), ('MZ', 0.0
3), ('SN', 0.03), ('AU', 0.03), ('DK', 0.03), ('LV', 0.03), ('KR', 0.03), ('IS',
 0.03), ('MA', 0.03), ('SA', 0.02), ('ID', 0.02), ('PK', 0.02), ('TW', 0.02)]
>>>
```

Obtención de indicadores de uso en RI

Adecuación y producción

- Instalación local para prueba y configuración de expresiones regulares
- Detección y solución de problemas o incompatibilidades
- Instalación en servidor de producción
- Importación de logs del período requerido, por ej. un año.
- Configuración de un evento cron para la importación automática, por ej., al momento de la rotación del archivo de log

Obtención de indicadores de uso en RI

Testeo

Agradecemos la colaboración de Nülan (UNMdP, Fac. Cs.Económicas - EPrints), Biblioteca Digital de UNCuyo (software desarrollo propio), RPsico (UNMdP, Fac. Psicología - DSpace) para el testeo de la aplicación

Obtención de indicadores de uso en RI

Licencia y descarga

Licencia Apache 2.0 (recomendada por GNU para pequeños programas)

Descarga desde el blog del Proyecto PICTO:
http://pictobonaerense.wordpress.com

Obtención de indicadores de uso en RI

Contenido

El archivo **indicadores.tar.gz** contiene los componentes de la aplicación Indicadores, el manual de usuario y una ponencia sobre el tema.

Obtención de indicadores de uso en RI

FIN

Íconos: gentileza de FreePik.com

¡¡Muchas Gracias!!