

# ALGORITMO AUTÓMATA PARA LA INSTALACIÓN ESTRUCTURADA DSPACE EN UBUNTU, UTILIZADO EN LA IMPLEMENTACIÓN DEL REPOSITORIO INSTITUCIONAL DE LA UNIVERSIDAD TECNOLÓGICA DE PANAMÁ

HURIVIADES CALDERÓN HURIVIADES.CALDERON@UTP.AC.PA;

## RESUMEN

El presente trabajo está enfocado en mostrar la creación de un algoritmo autónomo estable para la instalación personalizable de la plataforma DSpace aplicando el método *scripting* en cuatros módulos escrito en Bash. Su objetivo fue minimizar el tiempo de instalación y eliminar errores surgidos en la instalación de paquetes, permisos a carpetas y acceso a los diferentes roles de usuarios en el SO Ubuntu al momento de la instalación. Se muestran los pasos involucrados en el desarrollo de los módulos que conforma al algoritmo donde también se permite crear variantes o mejoras al código, ya que los módulos están bajo licencia *Attribution-NonCommercial-ShareAlike 4.0 International*. Esta iniciativa surge de la necesidad de implementar una plataforma para la creación del Repositorio Institucional de la Universidad Tecnológica de Panamá cuyo objetivo era centralizar y visibilizar los trabajos de investigación, académico y administrativos realizados por docentes e investigadores de nuestra Institución. Se muestra un breve análisis de porque es necesario implementar Repositorios Institucionales como también algunas estadísticas a nivel mundial y centroamericano del uso de DSpace como Plataforma.

**PALABRAS CLAVE:** Repositorio, algoritmo, DSpace, plataformas digitales, Ubuntu

## INTRODUCCIÓN

En el año 2014 se realizó en Buenos Aires Argentina un estudio sobre Indicadores de Acceso Abierto y comunicaciones Académicas en América Latina, éste estudio fue realizado por la UNESCO en conjunto con organizaciones como Redalyc, CLACSO, Scielo, FLACSO (Brasil) y PKP (*Public Knowledge Project*). Según el informe, en las regiones en desarrollo las investigaciones están financiadas exclusivamente por los gobiernos y gran parte de los resultados de las investigaciones se publican solamente en revistas académicas locales o regionales. La razón, se debe al uso de indicadores académicos llamados “índice de impacto internacionales” principalmente utilizados por directorios como *Web of Science* de Thomson-Reuters (WoS) y Scopus de Elsevier para evaluar el impacto de las revistas y publicaciones a nivel mundial. (Alperín y Babini, 2014).

Según datos de SCOPUS el total de revistas indexadas hasta el 2015 fue de 33,456, de este listado solo el 1.4% son Revistas de Latinoamérica y solo 7.4% del total son de acceso abierto. Si comparamos el número de Revistas latinoamericanas indexadas en SCOPUS con otra plataforma de indexación como Latindex, que contiene 12,990 revistas de Latinoamérica, España y Portugal, cerca del 43% son de Latinoamérica (Latindex, 2015) por lo que es evidente que la diferencia de publicaciones puede verse limitada a nivel Internacional en plataformas ser indexadas en SCOPUS y WoS.

Para mediar en este problema una de las propuestas surge bajo la cobertura de la declaración de Butapest de acceso abierto en el 2002, donde se da el apoyo de visibilizar los resultados de la investigación de la comunidad científica mundial en acceso abierto. En esta declaración se presentan dos alternativas, acceso abierto dorado, que es el acceso abierto a través de revistas científicas digitales con revisión por pares y el acceso abierto verde, que es el acceso abierto a través de repositorios, mediante archivado de artículos científicos de acceso abierto en formato electrónico (Berrocoso, 2013).

Tomando como referencia la alternativa de acceso abierto dorado, la Universidad Tecnológica de Panamá (UTP) da inicio en enero de 2016 la implementación del Portal de Revistas Académicas UTP utilizando la plataforma OJS (*Open Journal System*) para mostrar las revistas en formato digital en acceso abierto, este trabajo se realizó a 50 volúmenes de 8 Revistas que inicialmente estaban en formato impreso y dieron como resultado la catalogación, estructuración y digitalización de 900 artículos transformados en formato PDF y html y donde posteriormente se realizó un trabajo de indexación en *Google Scholar* (Murillo, 2016).

En marzo de ese mismo año, se dio inicio al proyecto UTP-Ridda2 en la UTP, con el objetivo de crear el Repositorio Institucional (RI) de Acceso abierto para centralizar toda la información científica y académica de la UTP, para tal fin se hizo un análisis de las Plataformas para implementar RI más utilizadas a nivel mundial según Opendoar.

OpenDoar es un directorio internacional de repositorios académicos de acceso abierto desarrollado por la Universidad de Nottingham del Reino Unido que hasta mayo del 2017 contaba con 3344 repositorios institucionales (OpenDoar, 2017). En datos estadísticos de esta plataforma el 44.1% de las instituciones utiliza DSpace, 13.7% Eprints, 4.8% Digital Commons.

En el caso de América central que cuenta con 19 RI de Países como Costa Rica, El Salvador, Honduras, Nicaragua y Guatemala donde el 63.2% utiliza DSpace, el 26.3% Eprints y 10.6% otros.

Otra referencia de esta plataforma es que en el 2010 se realizó un estudio de 11 plataformas de Repositorios y diferentes características, 6 plataformas eran OpenSource (DSpace, EPrints, Fedora, Islandora, IntraLibrary y Zenty) y 5 comerciales, dentro de las características que se analizaron estaban: soporte, instalación, requerimientos, escalabilidad, almacenamiento, sostenibilidad e interoperabilidad, actualización, migración y desempeño, mostrando que Dspace era una de las plataformas que mejor cumplía con estas características (Project, 2010).

Con estos resultados a nivel Mundial y Centroamericanos, la Plataforma que seleccionamos para el RI UTP fue DSpace, enfocados también en que existe un proyecto para la creación de un Repositorio Centroamericano SIIDCA-CSUCA ([www.repositoriosiidca.csuca.org](http://www.repositoriosiidca.csuca.org)) donde las Universidades que se vincularán a este repositorio cuentan con esta plataforma. DSpace, es un software de código abierto diseñado por el *Massachusetts Institute of Technology* (MIT) y los laboratorios de HP para gestionar repositorios de documentos digitales (textuales, audio, vídeo, etc.), facilitando su depósito, organizándolos en comunidades y colecciones, asignándoles metadatos y permitiendo su difusión a recolectores internacionales con objetivo de darle una mayor visibilidad a la producción científica y académica (Rodríguez y Sulé, 2008).

### Problemática

Aunque DSpace cuenta con los elementos para ser una óptima plataforma RI, una de sus desventajas es el proceso de instalación, debido a las numerosas configuraciones y dependencias requeridas como: Maven, OpenJDK, Tomcat y entre otras. A esto se suma que en el proceso se requiere cambiar de diferentes usuarios (*postgres*, *dspace* y *root*) múltiples veces, con el objeto de asignar roles y permisos correspondiente a las carpetas involucradas y modificar los archivos de configuración (*postgresql.conf*, *server.xml*, *tomcat7*, *environment*, *profile* y entre otras) en el servidor.

### OBJETIVO

Crear un algoritmo autónomo que permita realizar la instalación de DSpace 6.x en servidores Ubuntu 16.10 verificando cada componente necesario, modificando los permisos y mostrar la interface completa de Mirage 2 funcional con los roles asignados.

### MÉTODO

Para la realización de este estudio, se ha optado por la aplicación del método scripting, debido a que ofrece un entorno ideal para la configuración, compilación, instalación y personalización del código fuente de DSpace, usando este método se desarrolló cuatro módulos escrito en Bash. Como seguimiento de esta actividad, se mencionarán los módulos y sus tareas dentro de la ejecución estable de DSpace en el servidor:

#### PM-Prerequisites

Este primer módulo tiene como propósito de verificar todas las dependencias instaladas y sus versiones para DSpace en el sistema; de lo contrario el script utiliza el repositorio APT para obtener las dependencias faltantes o actualizar las versiones obsoletas. Como complemento a esta investigación, mencionaremos las dependencias y sus versiones recomendadas para DSpace 6.x (The DSpace Developer Team, 2017b): PostgreSQL 9.5.7, OpenJDK 8, Apache Tomcat 7.0.68, Apache Ant 1.9.6, Apache Maven 3.3.9, npm 3.5.2, Apache HTTP 2.4.18, nodeJS 4.2.6, ufw 0.35 y git 2.7.4.

Además, este módulo se encarga de configurar el firewall (iptables) del servidor, utilizando a UFW como front-end de iptables, para aplicar una regla por defecto de denegar todas las conexiones entrantes, exceptos los puertos 80, 443, 8080 y 8443, y de permitir todas las conexiones salientes de las aplicaciones *webs*.

#### P-A1

Este segundo módulo se encarga de compilar e instalar el DSpace en el servidor a través de tres funciones. La primera función, realiza la creación de la base de datos cifrada en PostgreSQL y un usuario propietario de la misma; la segunda función, habilita las conexiones TCP/IP de la base de datos utilizado un conector tipo JDBC y la tercera función, inserta las credenciales de la base de datos al archivo de configuración del DSpace para su posterior compilación. También, este módulo incluye por defecto el tema *web* Mirage 2, con el objeto de integrar un diseño web adaptable para los diversos dispositivos (computadoras, smartphones y tabletas) y sus resoluciones.

#### P-A2

Este tercer módulo define las variables de entorno utilizados en el servidor, por ejemplo: asignar el espacio de memoria en tiempo de ejecución (Java heap) de las aplicaciones en la máquina virtual de java (JVM), exportar la variable java (JAVA\_HOME) a los usuarios, adaptar el conector 8080 para el tráfico de los usuarios al DSpace, habilitar el puerto 8009 para el Apache JServ Protocol (AJP/Proxy) y enlazar

(*symbolic link*) las aplicaciones *webs* en modo de producción (*production*) al tomcat.

Adicionalmente, este módulo permite ejecutar DSpace localmente en el servidor.

### P-A3

Este último módulo despliega (deployment) las aplicaciones webs del DSpace (xmlui, oai y solr) hacia la internet. Para ello, se requiere un dominio web (DNS) apuntando al IP público, agregar el dominio al archivo de configuración del DSpace y crear un virtual host con los parámetros de configuración ProxyPass / ProxyPassReverse apuntando al IP local del servidor.

Cabe destacar, que los módulos se encuentran en la versión candidata a definitiva (RC) disponible solamente para la Distribución Ubuntu LTS 16 o superior. Por otro lado, se utilizó como base la documentación proporcionada por los desarrolladores de DSpace (The DSpace Developer Team, 2017a) para el desarrollo de las instrucciones algorítmicas de los módulos.

## RESULTADOS

Para evaluar la eficiencia de la ejecución de los módulos desarrollados, se utilizó como métrica la prueba de carga de 50 usuarios virtuales activos, localizado en Estados Unidos, y sus solicitudes durante 5 minutos consecutivo. El rango de aceptación para métrica sería menos de 1.5 segundos (Wang, Krishnamurthy, y Wetherall, 2016), cuyo resultado arrojó un tiempo promedio de carga de 550.02 ms ~ 761.59 ms. Otra tarea prioridad es asegurar la estabilidad del servidor en el consumo de memoria asignada, 2048 MB como límite, dentro de un rango de tolerancia del 7% al 15%; de lo contrario se provocaría un desbordamiento de memoria en el servidor.

Un hallazgo importante obtenido fue que la primera versión del script no se tomaba en cuenta un límite en el JVM y este provocaba que el DSpace saturara la memoria RAM en un periodo aproximado de 7 días causando la detención forzada del repositorio. Por otra parte, se debe asignar tareas automáticas al usuario encargado del DSpace para optimizar diariamente el OAI y eliminar el excedente de memoria no asignada para tomcat. No obstante, estos problemas fueron superados en la versión RC de los módulos. Como resultado de esta investigación, compartiremos el URL ([www.ridda2.utp.ac.pa](http://www.ridda2.utp.ac.pa)) que aloja el Repositorio Institucional de la Universidad Tecnológica de Panamá.

## CONCLUSIONES

Con el desarrollo del presente estudio se logró completar los objetivos propuestos, obtener un algoritmo autónomo estable para la instalación personalizable del DSpace, según las necesidades del entorno académico. Al mismo tiempo, se

logró demostrar todos los pasos involucrados en el desarrollo de los módulos que conforma al algoritmo. Además, se permite crear variantes o mejoras al código, ya que los módulos están bajo licencia *Attribution-NonCommercial-ShareAlike 4.0 International* (CC BY-NC-SA 4.0).

Como trabajos futuros se tiene contemplado el soporte para CentOS y sería de interés implementar otros módulos adicionales como, por ejemplo: personalización de la metadata en Mirage 2, crear APIs propias para manipular la metadata procedente del DSpace y adaptar el módulo P-A2 para asignar dinámicamente el espacio de memoria requerido para Java en base a los recursos disponible en el servidor.

## REFERENCIAS BIBLIOGRÁFICAS

- Berrocoso, J. (2013). El acceso abierto al conocimiento científico. *Reuni+D*, 56. Doi: <https://doi.org/http://hdl.handle.net/2445/36335>
- Murillo, D. (2016). Análisis de la producción científica y académica en revistas institucionales para mejorar la visibilidad a nivel nacional e Internacional a través de una plataforma tecnológica. *Educación y Tecnología*, 2-3. Octaedro.
- Alperín, J. Babini, G. (2014). *Indicadores de acceso abierto y comunicaciones académicas en América Latina*. Recuperado de: [http://biblioteca.clacso.edu.ar/clacso/se/20141217052547/Indicadores\\_de\\_acceso\\_abierto.pdf](http://biblioteca.clacso.edu.ar/clacso/se/20141217052547/Indicadores_de_acceso_abierto.pdf)
- Latindex. (2015). Listado de Revistas Científicas por Región.
- OpenDoar. (2017). Repositorios Centroamericanos.
- Project, R. S. (2010). Repository software survey.
- Rodríguez, J. y Sulé, A. (2008). DSpace: un manual específico para gestores de la información y la documentación. *BiD Textos Universitaris de Biblioteconomia I Documentaci*, 1-15.
- The DSpace Developer Team. (2017a). DSpace 6.x Documentation, (July).
- The DSpace Developer Team. (2017b). Prerequisite Software. Retrieved from <https://wiki.duraspace.org/display/DSDOC6x/Installing+DSpace#InstallingDSpace-PrerequisiteSoftware>
- Wang, X., Krishnamurthy, A., y Wetherall, D. (2016). Speeding up Web Page Loads with Shandian. *Ndsi*, 15. Recuperado de: <https://www.usenix.org/conference/nsdi16/technical-sessions/presentation/wang>