

Webbot de Datos Bibliométricos (WDB)

*De Domingo Gonzalo, Guastadisegni Federico, Rodríguez Nahuel
Licenciatura en Sistemas – Departamento de Desarrollo Productivo y Tecnológico
Universidad Nacional de Lanús.*

Resumen-El presente documento trata sobre el proceso de exploración de conceptos realizada para el proyecto Webbot de Datos Bibliométricos, perteneciente a la cátedra Proyecto de Software de la carrera Licenciatura en sistemas, Universidad Nacional de Lanús.

Palabras Clave – Requerimientos de usuario, necesidades.

III. Procesos de predesarrollo

A. Introducción

El planteo del sistema surge a partir de la necesidad de crear un artefacto web que permita la recolección de información acerca de publicaciones artículos científicos que contengan los indicadores bibliométricos solicitados por el cliente.

Un Webbot, es una herramienta diseñada para poder recorrer ciertos dominios de internet en base a una o varias palabras clave, para posteriormente devolver esa información al servidor.

En este documento vamos a describir cuáles son los requisitos funcionales que consideramos necesarios para el desarrollo de nuestro artefacto software. Para llegar a ello, previamente realizamos un cuestionario con nuestras inquietudes a nuestros clientes (profesores de la cátedra), así como también la realización de preguntas en clase.

B. Requisitos de software

En esta sección se enunciarán los distintos tipos de requerimientos del artefacto software a desarrollar.

Hay 3 (tres) tipos de proyectos aplicables:

- Para las páginas “caóticas” (aquellas que no presentan un orden para la información, que no presentan una limitación entre los distintos elementos que componen la página, como el título, copete, texto, imágenes, etc.) se considera un tiempo de trabajo de 5 a 10 días.
- Para las páginas relativamente organizadas se considera un tiempo de trabajo de 3 a 5 días.
- Para las páginas correctamente organizadas y separadas en secciones adecuadas, se considera un tiempo de trabajo de 3 días.

C. Requerimientos del Usuario

Se requiere el desarrollo de un artefacto capaz de contabilizar los indicadores solicitados.

En esta sección se va realizar un análisis de la situación actual y las necesidades que llevan al desarrollo del artefacto software. Se contemplan diferentes soluciones (ver punto 4 más adelante) y se propone la alternativa elegida de común acuerdo con el cliente.

A) Proceso de exploración de conceptos

A continuación se detallan los diferentes procesos referidos a este tema.

1. Situación Actual

Actualmente los Web Bots (WB), o Web Crawlers son utilizados ampliamente para el indexado de paginas web. Esto es útil para paginas que utilizan motores de búsqueda, y que requieren obtener información relacionada a un tema o palabra en específico de una forma rápida, confiable y ordenada.

2. Dominio del problema

El objetivo del proyecto es elaborar un Web Bot (WB), dirigido a recolectar Datos Bibliométricos, almacenar esa información e indexarla con la finalidad de ser utilizada posteriormente. Se utilizarán hasta un máximo de 10 páginas con información bibliométrica, con un formato ya definido.

3. Informe preliminar de necesidades

Se han detectado las siguientes necesidades:

- Entender el objetivo y las limitaciones que tiene un Web Bot (WB) desarrollado en un ambiente universitario.
- Desarrollar la aplicación utilizando el lenguaje Python versión 2.7.X.
- El sistema es un intermediario, cuya información recolectada será utilizada por un usuario.
- El sistema basará sus búsquedas en una "plantilla" predeterminada, que será creada a partir de tags presentes en 5 o 10 páginas que contengan material bibliométrico, en las cuales estarán presentes los indicadores propuestos por el cliente.
- Dependiendo del tipo de página, se estima un tiempo aproximado de desarrollo.

4. Soluciones Alternativas

El software puede ser desarrollado sin la estructura de un patrón de diseño conocido, o usando algún patrón como por ejemplo MVC (No utilizaremos Vistas, se construirá una API de testeo).

Pueden utilizarse distintos paradigmas como por ejemplo el Orientado a Objetos, entre otros.

El lenguaje Python es un requisito para el desarrollo del artefacto, por lo tanto se analizaron distintos Frameworks y librerías correspondientes a dicho lenguaje que podrían llegar a ser útiles para el desarrollo del artefacto, como por ejemplo: Django, Scrapy, entre otros.

5. Solución Recomendada

Luego de conversar con el cliente y explorar las distintas alternativas, se decidió de común acuerdo utilizar un patrón de diseño que incluya modelo y controlador, y una API de "testeo" para nuestro uso en la fase de pruebas.

Por otro lado, se observa en el Framework Scrapy un buen soporte para el desarrollo de la aplicación, ofreciendo además las herramientas necesarias para la misma, por lo que se decidió junto al cliente que el mismo sería el más conveniente para el desarrollo de nuestro proyecto.

Se utilizará el ORM (mapeo objeto-relacional) de Django y se integrará junto a Scrapy.

Versión	Detalle de la versión		
	Descripción	Autor	Fecha
1.00.001	Primera entrega.	GdeDomingo FGuastadisegni NRodríguez	14/08/15
1.00.002	Se dividió la documentación conjunta en varios documentos individuales correspondientes a los diferentes procesos de la IEEE/1074-1989.	NRodríguez	25/09/15
1.00.003	Se modificó el contenido en función a lo conversado con el cliente (se incluyó en la sección correspondiente la mención a : API de testeo y obtención de información rápida, confiable y ordenada) y se hicieron correcciones menores en otros apartados.	NRodríguez	25/09/15
1.00.004	Se modificó la sección “dominio del problema”. Se incluye ahora la cantidad máxima de páginas a ser utilizadas por el Webbot. Además, se separó en una oración el tratamiento que se le hace a la información (recolectar, almacenar e indexar, para luego mostrar) y en otra desde dónde se obtiene dicha información.	NRodríguez	02/10/15