

# Módulo de Estadísticas para ARCA

◆ David Pérez Redondo, José Ignacio Pérez Alcocer

## Resumen

Como en cualquier herramienta/medio de distribución o difusión de contenidos, en ARCA se hace necesario conocer datos de acceso a los mismos. Con esta idea nace el módulo de estadísticas para permitir saber qué vídeos y eventos son los más visitados y quién y cuándo consulta dichos contenidos.

## Summary

Since in any tool / way of distribution or diffusion of contents, in ARK it becomes necessary to know information of access to the same ones. With this idea the module of statistics is born to allow to know that videos and events are most visited and, who and when it consults the above mentioned contents.

## 1. Introducción

La creación de un módulo de estadísticas en ARCA se plantea a raíz de tener la necesidad de saber cuál es la demanda que tienen los contenidos multimedia publicados en ARCA. Lo que se pretende con este módulo es ofrecer al usuario una idea del alcance y el tipo de audiencia que tienen los contenidos publicados dentro del portal. Estos contenidos pueden ser de tres tipos: evento en directo, vídeo bajo demanda y podcast/vodcast. Si bien es cierto que estas estadísticas no van a sustituir a las obtenidas por los distintos servidores de streaming, este módulo ayudará a las instituciones que no dispongan de un sistema propio de estadísticas, y para las que sí dispongan de dicho sistema, complementará la información obtenida, puesto que reflejará datos sobre contenidos de ARCA en los que no todos son puramente streaming.

## 2. Arquitectura

Para el desarrollo del módulo de estadísticas hemos utilizado una plataforma Debian Sarge con Apache2, el código está implementado en PHP5, y el servidor de base de datos empleada es MySQL 5. En cuanto a las librerías gráficas hemos evaluado cuatro distintas: JGraph, libchart, GnuPlot y Artichow, y finalmente elegimos utilizar las librerías de Artichow ([www.artichow.org](http://www.artichow.org)), porque nos permitían trabajar con varias variables en las mismas gráficas, el diseño del aspecto de las gráficas es muy flexible y se licencia como creative commons (by-nc-sa), lo que nos permite distribuir y hacer obras derivadas para uso no comercial.

Cada vez que un usuario realiza una consulta sobre un ítem (entendiendo por ítem el contenedor del anuncio de un VoD, podcast o evento en directo), el portal ARCA anotará esa consulta como un registro en una tabla de la BB.DD.; posteriormente el módulo de estadísticas recogerá esa información y la procesará para mostrar gráficamente todas las consultas realizadas, ordenadas por meses, canales, sistema de clasificación empleado, etc.

El procesamiento de estadísticas consiste en tres fases: recogida de datos, procesado de datos y visualización de datos.



Las estadísticas ofrecerán al usuario información sobre el alcance y la audiencia que tienen los contenidos del portal de ARCA



Este módulo ayudará a las instituciones que no dispongan de un sistema propio de estadísticas

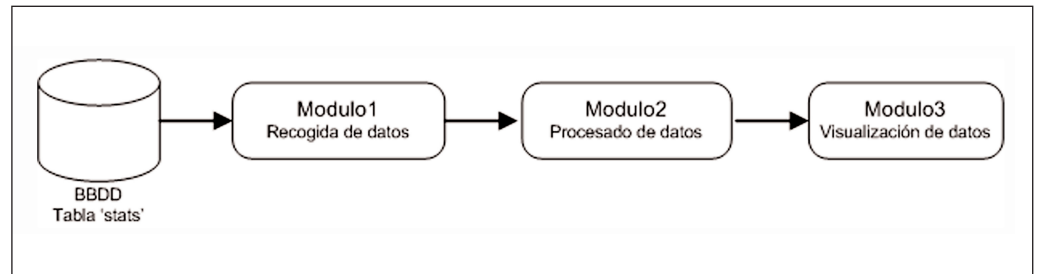


Figura 1. Procesamiento de estadísticas.

El modo de procesamiento en tres fases permite la creación de plantillas de estadísticas

En la primera fase, “Recogida de datos”, se acumularán los registros de la tabla de estadísticas en tablas de datos históricos, y se eliminarán de la misma. Esto evitará que la tabla de estadísticas tome dimensiones de difícil manejo. Este módulo se puede ejecutar por separado, de manera que la recogida de datos se podrá realizar en cualquier momento. Justo antes de que ARCA realice la actualización de los RSS de los distintos canales, se ejecutará el script de procesamiento de estadísticas, de esta manera aseguraremos una información estadística actualizada diariamente.

Las fases de procesado y visualización de datos se ejecutarán vía web, cuando el usuario realice la petición de unas estadísticas concretas, p. ej.: estadísticas por canal del mes de mayo de 2007.

En la segunda fase, “Procesado de datos”, se realizan los cálculos necesarios para que los componentes de visuales puedan obtener los datos necesarios para la visualización de las estadísticas correspondientes.

La tercera fase, “Visualización de datos”, se encarga de mostrar gráficamente los datos procesados. Este módulo es un conjunto de componentes visuales o widgets que, al hacer uso de las características propias del paradigma de la orientación a objetos, son fácilmente reutilizables en las distintas plantillas.

La plantilla de estadísticas por canal muestra el ranking de vídeos más visitados y las categorías más consultadas

Este modo de procesamiento en tres fases nos permite la creación de plantillas de estadísticas que favorecen la reutilización de componentes, y facilita la tarea de añadir nuevos componentes y funcionalidades.

### 3. Estadísticas en la actualidad

En estos momentos existe una primera versión en pruebas en el portal ARCA de la UC3M. En esta versión está disponible la plantilla de estadísticas por canal, donde se puede obtener las estadísticas de visitas a los contenidos de un canal ordenados por meses. Se puede ver también el ranking de los vídeos más visitados y el de las directrices o categorías más visitadas.

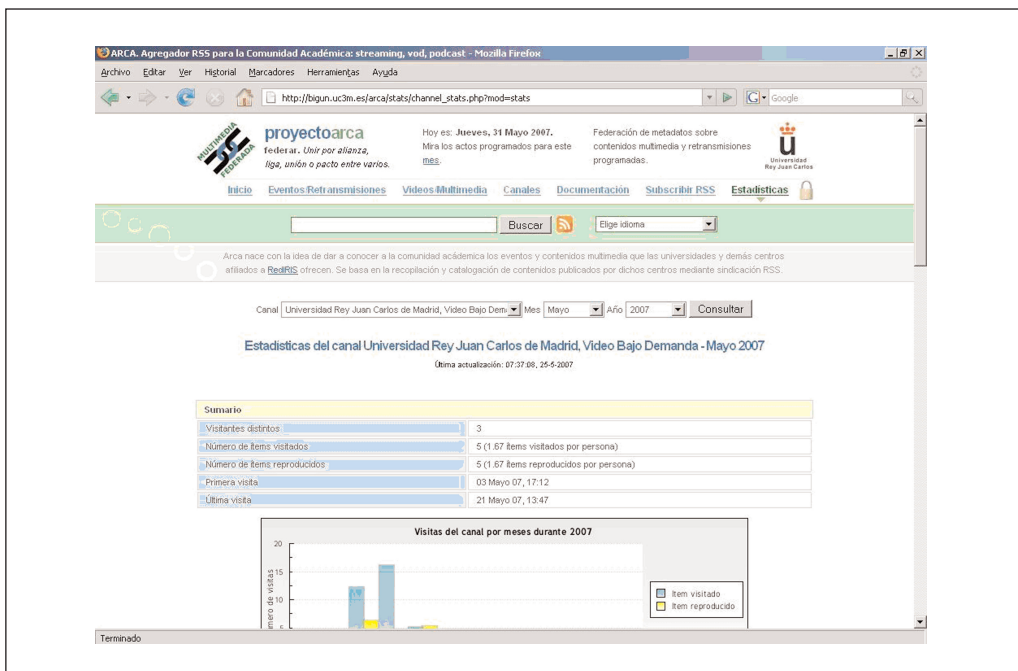


Figura 2. Plantilla de canal - Sumario.

La herramienta permite consultar las visitas del año y del mes

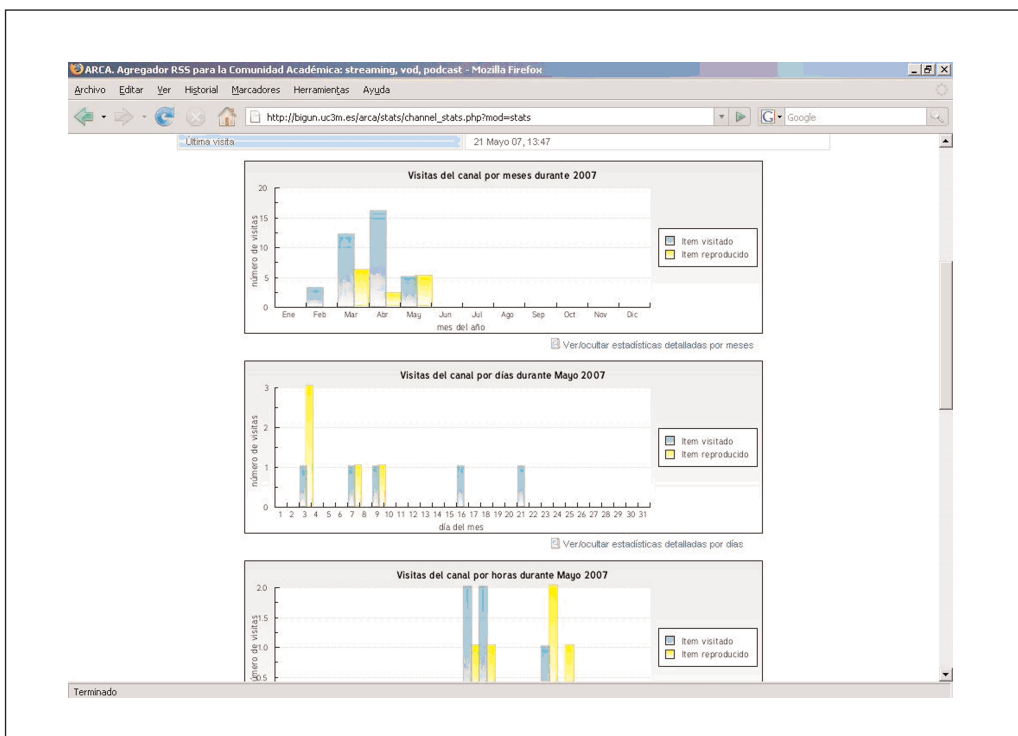


Figura 3. Plantilla de Canal- Visitas del año y del mes.



Está previsto realizar las plantillas de estadísticas por categorías y estadísticas generales de ARCA

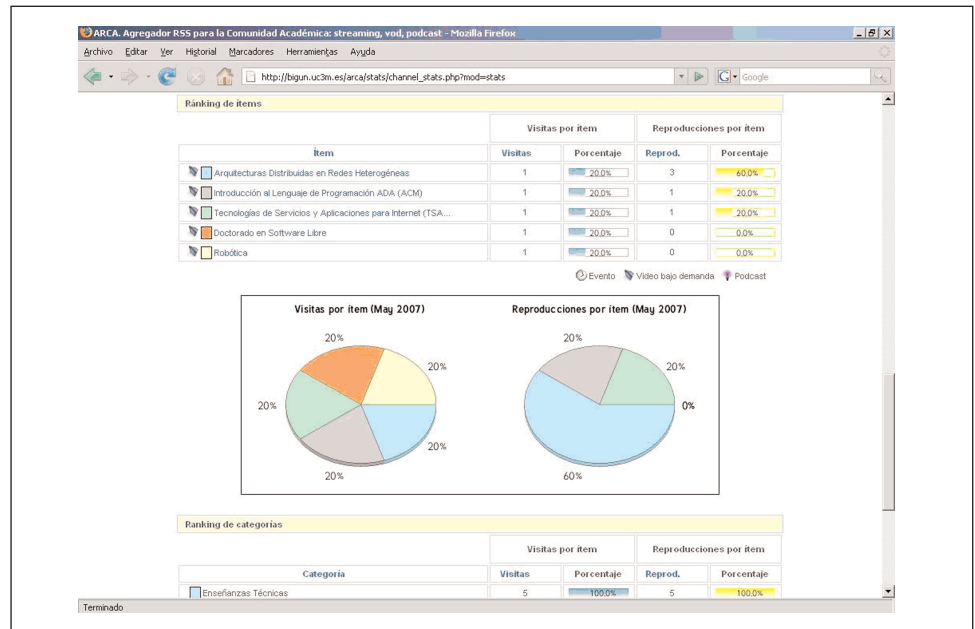


Figura 4. Plantilla de Canal- Ranking de ítems.

## 4. Líneas de trabajo futuras

Desde el grupo de Audiovisuales del Servicio de Infraestructuras Tecnológicas de la URJC, estamos trabajando para tener completamente operativa la plantilla de estadísticas por canal, añadiendo los ranking de dominios de procedencia de las visitas al canal, y los navegadores más utilizados por los visitantes.

Además, la previsión es realizar al menos dos plantillas más: estadísticas por categorías y estadísticas generales de ARCA, donde se recogerán los datos de consultas de contenidos de todos los canales.

## 5. Colaboración y agradecimientos

Queremos agradecer a la UC3M el esfuerzo realizado en el desarrollo de este portal, que permitirá dar a conocer nuestro trabajo más allá de la frontera de nuestras instituciones, así como, a todas las personas e instituciones que apoyan de una u otra manera este proyecto.

De la misma manera, animar a todo aquel que quiera aportar su granito de arena, que se ponga en contacto con RedIRIS, la UC3M ó la URJC para que podamos abordar, entre todos, los problemas que vayan surgiendo.

**David Pérez Redondo**

(david.perez@urjc.es)

sip:7052@urjc.es

**José Ignacio Pérez Alcocer**

(joseignacio.perez@urjc.es)

sip:8513@urjc.es