



4as Jornadas OS-REPOSITORIOS

“Políticas de promoción del acceso abierto”

Barcelona, 3 - 5 de marzo de 2010

Análisis de la frecuencia de carga en grandes Repositorios centrales y Repositorios institucionales



Luis-Millán González¹, Fernanda Peset², Antonia Ferrer², Rafael Aleixandre³

¹ Universidad de Valencia

² Universidad Politécnica de Valencia

³ Universidad de Valencia-CSIC

Financiación parcial de:



INTRODUCCIÓN

En la actual sociedad de la información, donde cada día a día se **multiplica la cantidad de datos almacenados** casi de forma exponencial, la minería de datos es una herramienta fundamental para analizarlos y explotarlos de forma eficaz para los objetivos de cualquier organización.

LÍMITES COGNITIVOS



INTRODUCCIÓN

La MINERÍA DE DATOS se define como la extracción no trivial de información implícita, previamente desconocida y potencialmente útil, a partir de datos.



Algunos sistemas que son sólo parcialmente conocidos producen una cantidad inmensa de datos; estos datos con frecuencia contienen valiosa información que puede resultar muy útil para determinados procesos

INTRODUCCIÓN

Depósitos institucionales / temáticos

- **Temáticos:**
 - Creados en torno a una disciplina
 - Aceptan materiales relacionados con la temática del depósito
 - Imposible que sean completos
 - No hay forma de obligar a los autores a remitir sus trabajos
- **Institucionales:**
 - Centrados en una organización: Universidad, Departamento, Instituto...), Sociedades Científicas
 - Contiene solamente documentos de los miembros de la institución
 - Es posible definir políticas para que los miembros añadan contenidos
 - Beneficios claros para los gestores:
 - Los beneficios del impacto de la investigación se perciben a nivel institucional
 - El sistema de incentivos de la ciencia está organizado en torno a las instituciones

INTRODUCCIÓN

OBJETIVOS

- Cuantificar el movimiento de archivos de los principales repositorios indexados en ROAR durante los años 2008 y 2009
- Comparar los movimientos realizados por los principales repositorios indexados en ROAR en función de su categoría (Centrales vs. Institucionales)
- Establecer la frecuencia de carga de los principales repositorios indexados en ROAR



MATERIAL Y MÉTODO

Welcome to the Registry of Open Access Repositories - Registry of Open Access Repositories - Mozilla Firefox

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

http://roar.eprints.org/

Most Visited Getting Started Latest Headlines

Registry of Open Access Repositories (ROAR)

Home Search Search Content Browse by Country Browse by Type Browse by Software Register in ROARMAP About

Login Create Account/Add Repository

Welcome to the Registry of Open Access Repositories

The aim of ROAR is to promote the development of open access by providing timely information about the growth and status of repositories throughout the world. Open access to research maximises research access and thereby also research impact, making research more productive and effective. [More information...](#)

Any Country Any Software Any Repository Type Sort by number of reco Search

Displaying results 1 to 20 of 1638.

1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | Next

Export 1638 results as Activity table Export

RSS 2.0 RSS 1.0 Atom

Graphical analysis Search Content

- Cumulative Deposits Per Day

PubMed Central (1617457 records) - 27 February 2001 [Record Details](#)

Terminado

Bandeja de entrada ... Microsoft PowerPoi... Welcome to the Reg...

9:16

MATERIAL Y MÉTODO

Diseño

Descriptivo comparativo

Población

Todos aquellos repositorios indexados en ROAR

Muestra

Se realizó un muestreo no probabilístico por conveniencia (n=50). Para establecer comparaciones se respetó una proporción alícuota entre ambos tipos de repositorios (Centrales vs. Institucionales)

Any Repository Type	Sort by number of
Any Repository Type	
Database/A&I Index (36)	
Demonstration (19)	
Learning and Teaching Objects (20)	
Other (228)	
Research Cross-Institutional (151)	
<u>Research Institutional or Departmental (931)</u>	
e-Journal/Publication (107)	
e-Theses (147)	

MATERIAL Y MÉTODO

Procedimiento General

Todos los datos del presente trabajo fueron descargados empleando las aplicaciones contenidas en la web (<http://roar.eprints.org/>). Para ello se seleccionaron los 50 repositorios que más movimiento tuvieron durante el año 2008 y 2009.

Concretamente se descargaron 25 repositorios centrales y 25 institucionales empleando la utilidad (Activity table) disponible en la aplicación informática del sitio web.

Todos los datos fueron almacenados en un disco duro para su posterior análisis.

MATERIAL Y MÉTODO

Análisis estadístico

Supuestos
previos

Kolmogorov-Smirnov
Levene

¿ NORMALIDAD ?
¿ HOMOGENEIDAD ?

Estadística
descriptiva

Valores absolutos / descriptivos de tendencia
central y dispersión

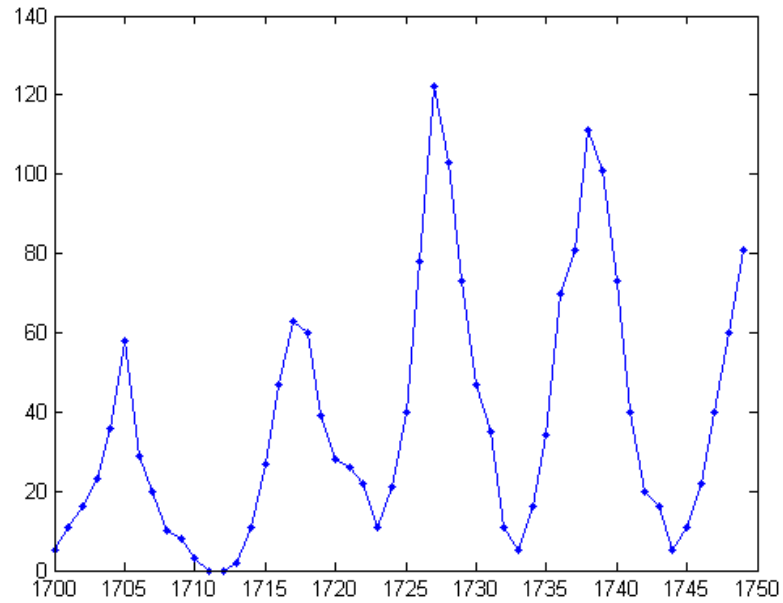
Estadística
inferencial

Prueba t de Student para muestras independientes

Análisis
Frecuencial

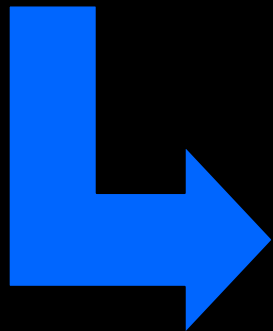
Fast Fourier Transformate

Una transformada de Fourier es una operación
matemática que transforma una señal de dominio de
tiempo a dominio de frecuencia y viceversa

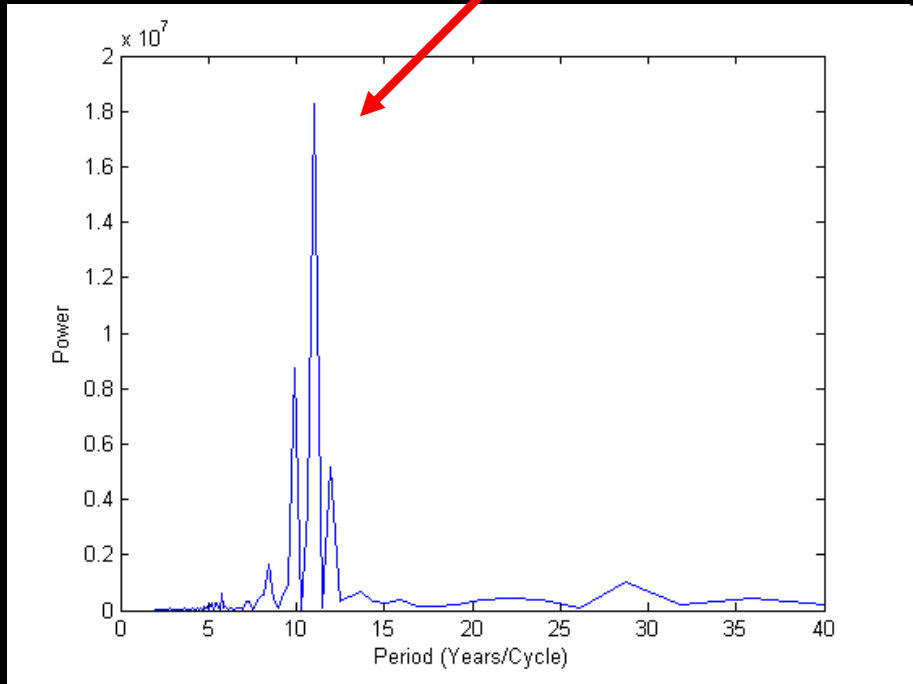


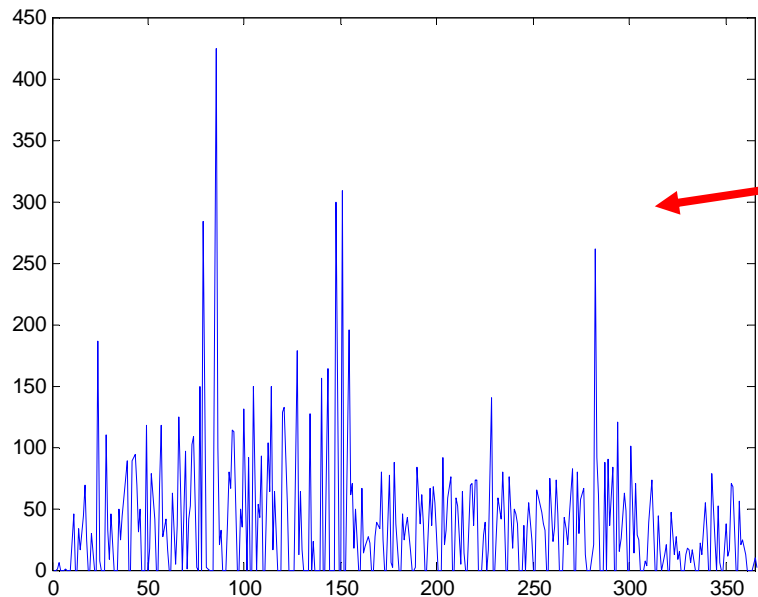
Manchas solares y tamaño

Frecuencia de aparición



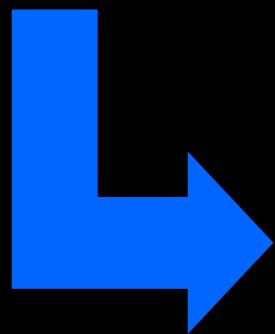
FFT



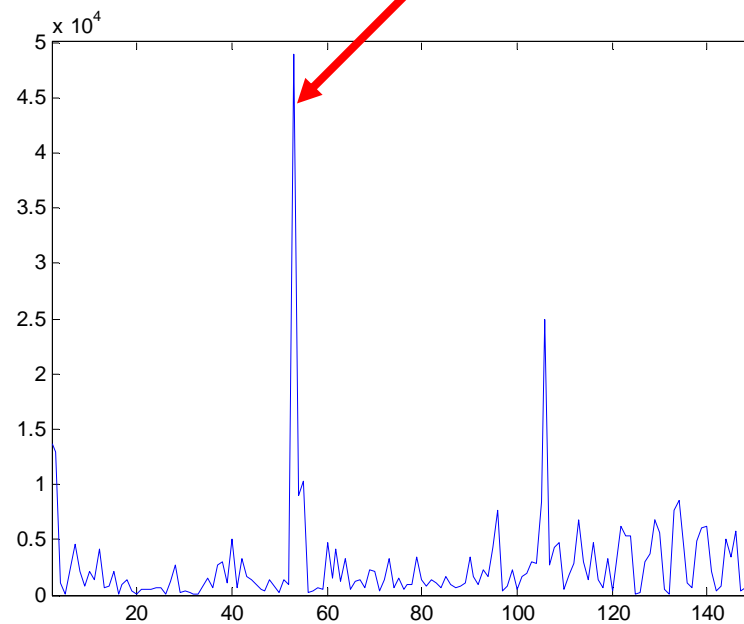


**Carga de documentos
Repositorio REPEC
Thomas Krichel**

**Frecuencia de
carga**



FFT



RESULTADOS

**TOTAL DE DOCUMENTOS CARGADOS DURANTE
LOS AÑOS 2008-2009**

3.001.548

**REPOSITORIO QUE MÁS CARGAS REALIZÓ
DURANTE
LOS AÑOS 2008-2009**

**898.140
(29,92%)**

→ soton southampton
(INS)

**REPOSITORIO QUE MÁS CARGAS REALIZÓ EN UN
DIA
LOS AÑOS 2008-2009**

**129.527
(4,32%)**

→ AGSPACE
(INS)

RESULTADOS

ESTADISTICOS DE LA MUESTRA (n=50)

	Total	Máximo carga en un día	CV	Promedio Semanal
Media	60.030,96	7.977,52	653,83	81,33
ET	(22804,63)	(3143,63)	(82,56)	(30,55)

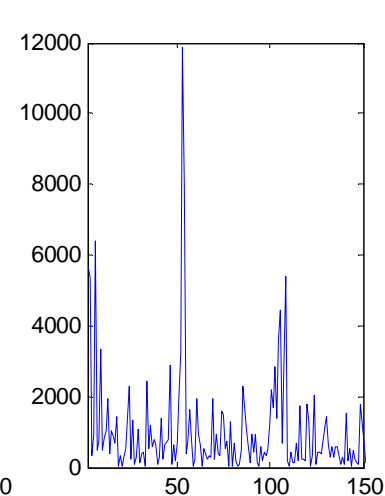
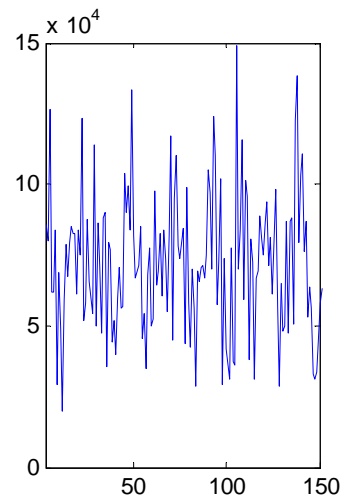
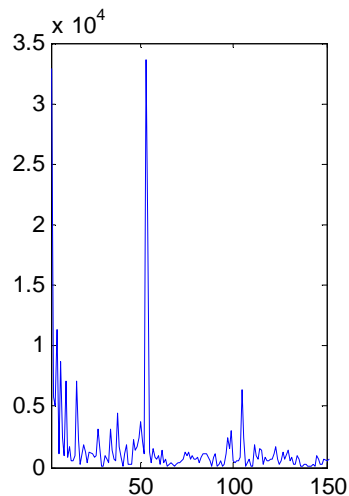
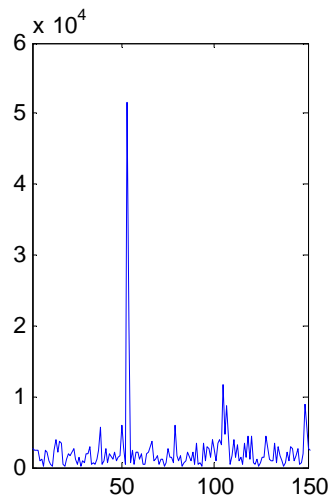
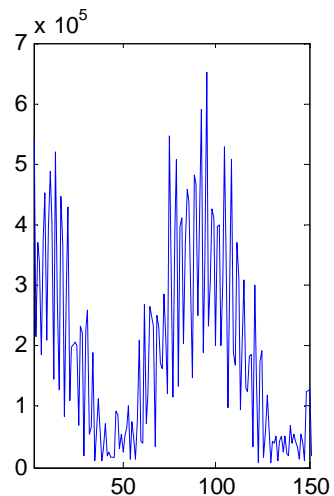
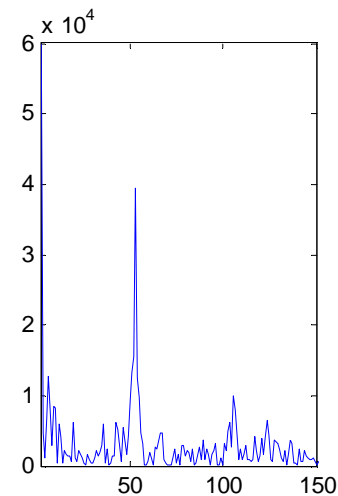
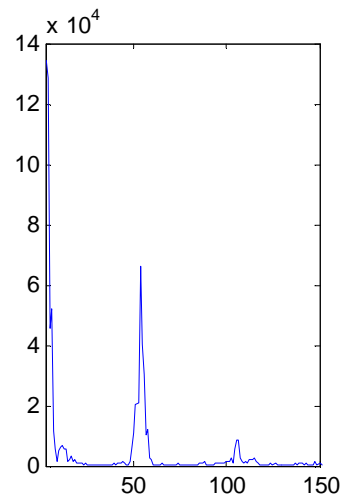
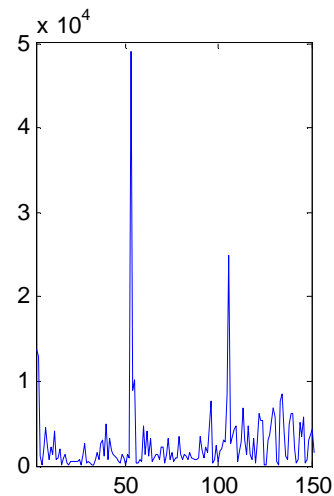
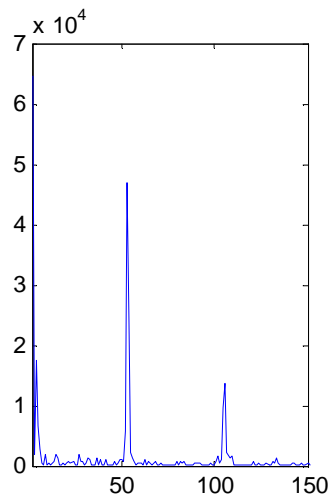
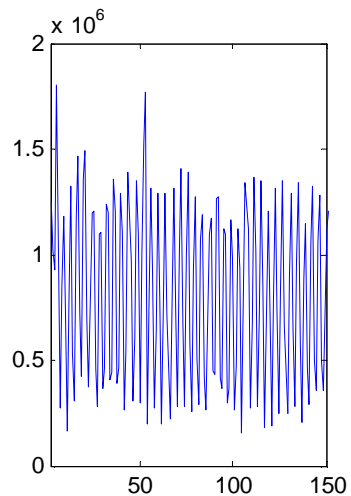
RESULTADOS

ESTADISTICOS EN FUNCIÓN DEL TIPO (CENTRALES; n=25, INSTITUCIONALES; n=25)

	grupo	Media	Error típ. de la media
Total	CENTRALES	28466,40	4630,95
	INSTITUCIONALES	91595,52	44934,06
Máximo de carga en un día	CENTRALES	4132,36	1039,79
	INSTITUCIONALES	11822,68	6167,65
CV	CENTRALES	567,84	82,85
	INSTITUCIONALES	739,82	142,67
Promedio Semanal	CENTRALES	39,02	6,35
	INSTITUCIONALES	123,64	60,17

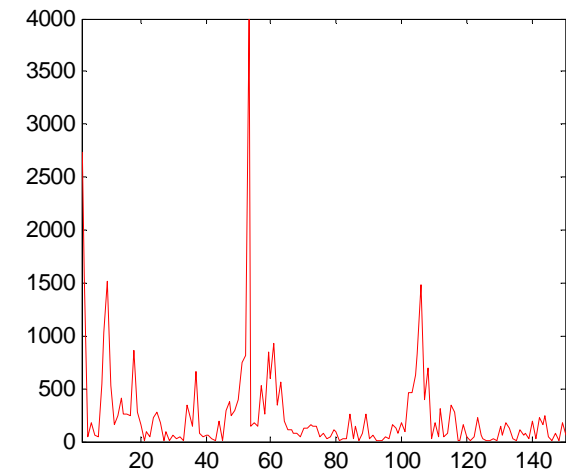
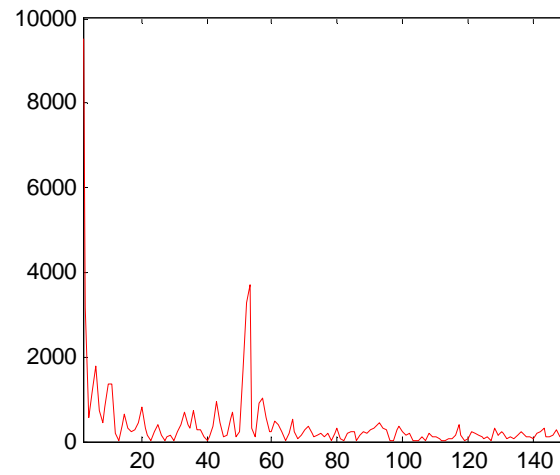
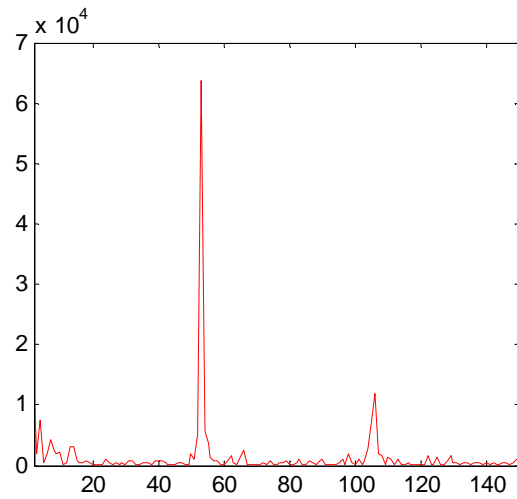
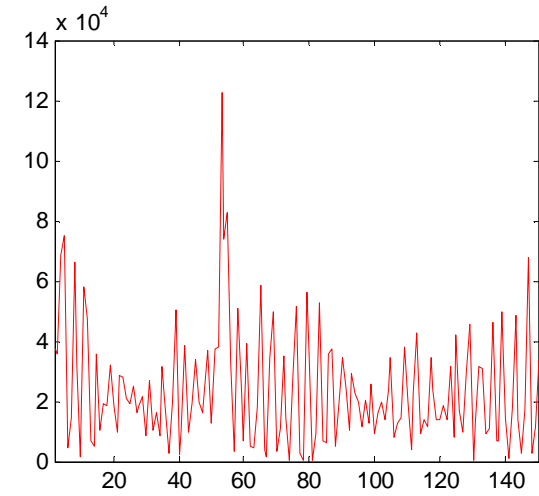
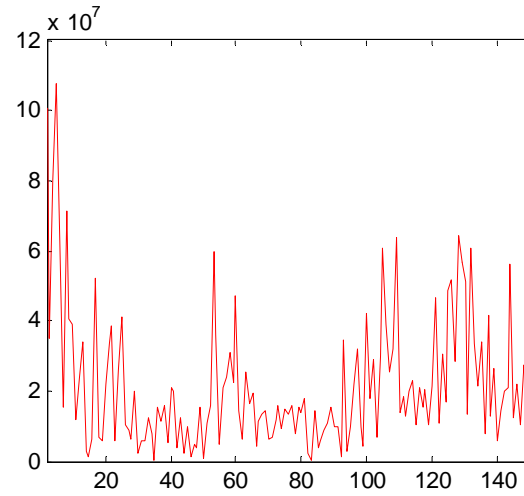
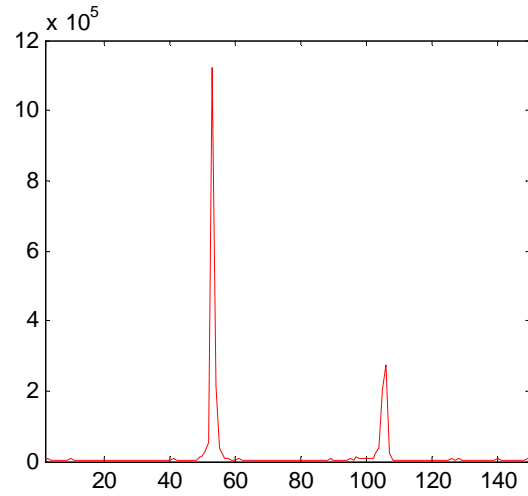
Resultados

CENTRALES



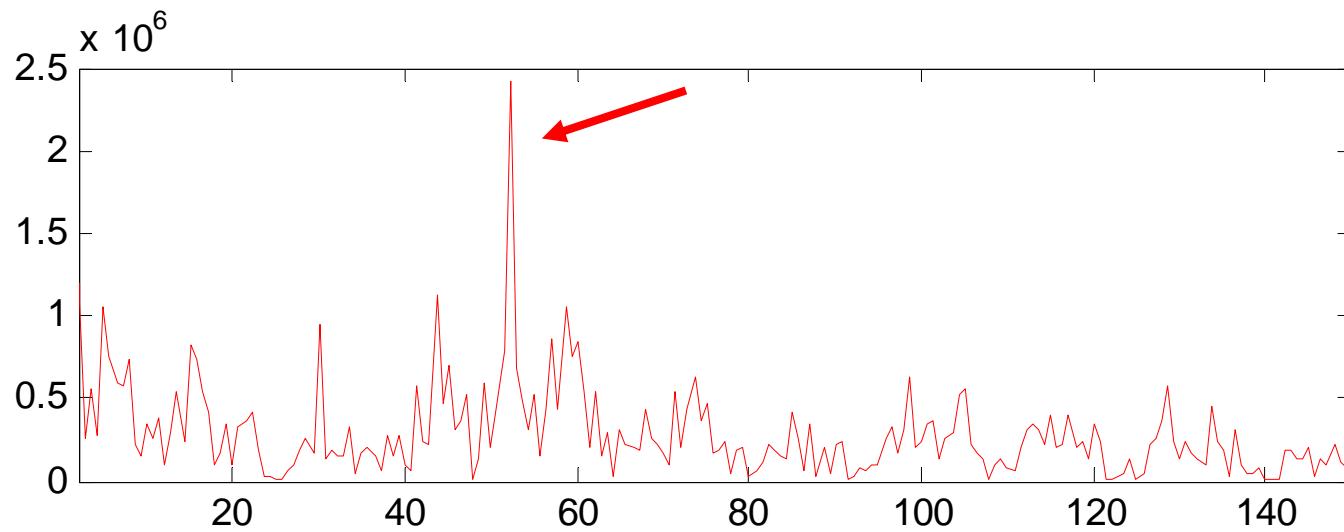
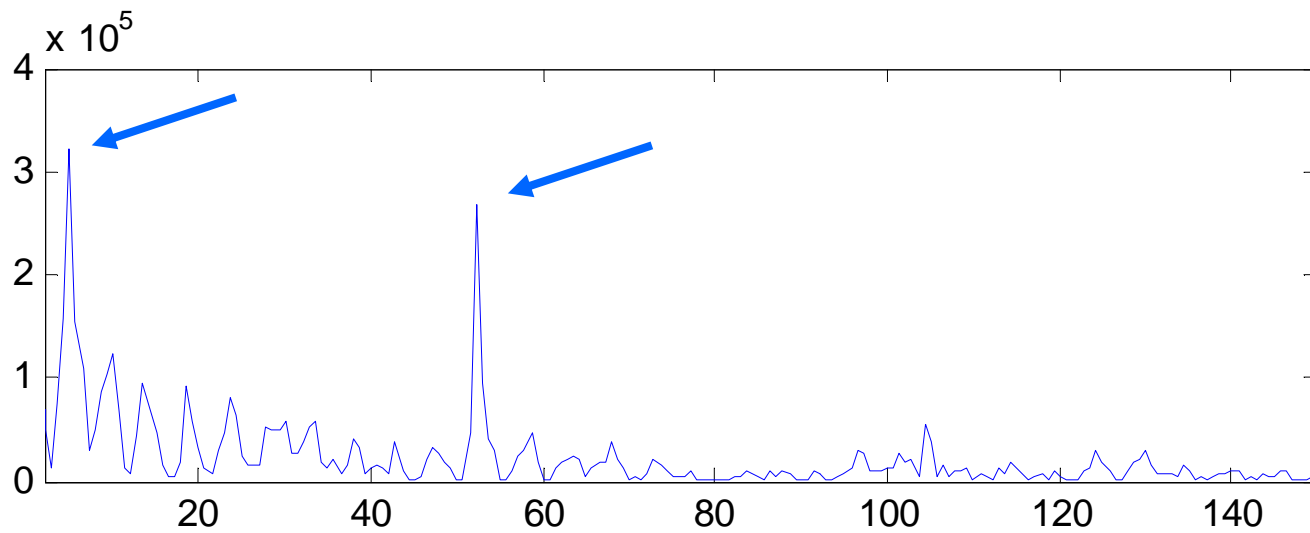
Resultados

INSTITUCIONALES



Resultados

CENTRALES / INSTITUCIONALES



CONCLUSIONES

- Los principales repositorios indexados en ROAR durante los años 2008 y 2009 tuvieron un número aproximado de carga de documentos entorno a los 3.000.000 de registros, llamando la atención que el 4% de estos documentos se cargaron en tan sólo un día por un repositorio
- Los repositorios Institucionales cargaron una mayor número de documentos que el grupo de comparación. No obstante, los repositorios Centrales muestran un patrón de carga más homogéneo y constante.
- La cuestión que cabe preguntarse es si las políticas de promoción que se aplican a los Institucionales, ¿los harán sostenibles en un futuro?
- Se encontró una frecuencia de carga de máxima amplitud entorno a los 50 días, teniendo este hallazgo claras aplicaciones tanto para las políticas de gestión de repositorios, como para las recomendaciones que se han de dirigir a los usuarios finales

BIBLIOGRAFÍA

Alonso J, Subirats I, Martínez Conde ML . *Informe APEI sobre acceso abierto*. APEI, 2008
<http://eprints.rclis.org/15107/1/informeapeiaccesoabierto.pdf>

Barton MR., Waters MM. Cómo crear un repositorio institucional. Manual LEADIRS II.. Mit Libraries, 2004-2005.

Decroos F, Dierckens K, Pollet V, Hugo R, Verweyen K. Spectral methods for detecting periodicity in library circulation data: a case study. *Information Processing & Management* 33 (1997) 393-403

Ferrer A, Moreno M, González LM, Peset F. Indicadores de rendimiento para acciones de acceso abierto, *Acceso Abierto: Métrica e indicadores*. FESABID, 2009
<http://www.slideshare.net/fesabid/antonia-ferrer-m-moreno-lm-gonzalez-y-f-peset-fesabid-fpeset-metrica?type=powerpoint>

Kao SC, Chang HC, Lin CH. Decision support for the academic library acquisition budget allocation via circulation database mining. *Information Processing & Management* 39 (2003) 133–147

Melero R, Barrueco JM. Acceso abierto y repositorios de documentos. SEDIC, 2007
http://www.sedic.es/autoformacion/acceso_abierto/1-Introducci%F3n-acceso-abierto.html

Peset F, Ferrer A. Implantación de Open Archives Initiative en España, *Information Research* 13 (dec. 2008) <http://informationr.net/ir/13-4/paper385.html>