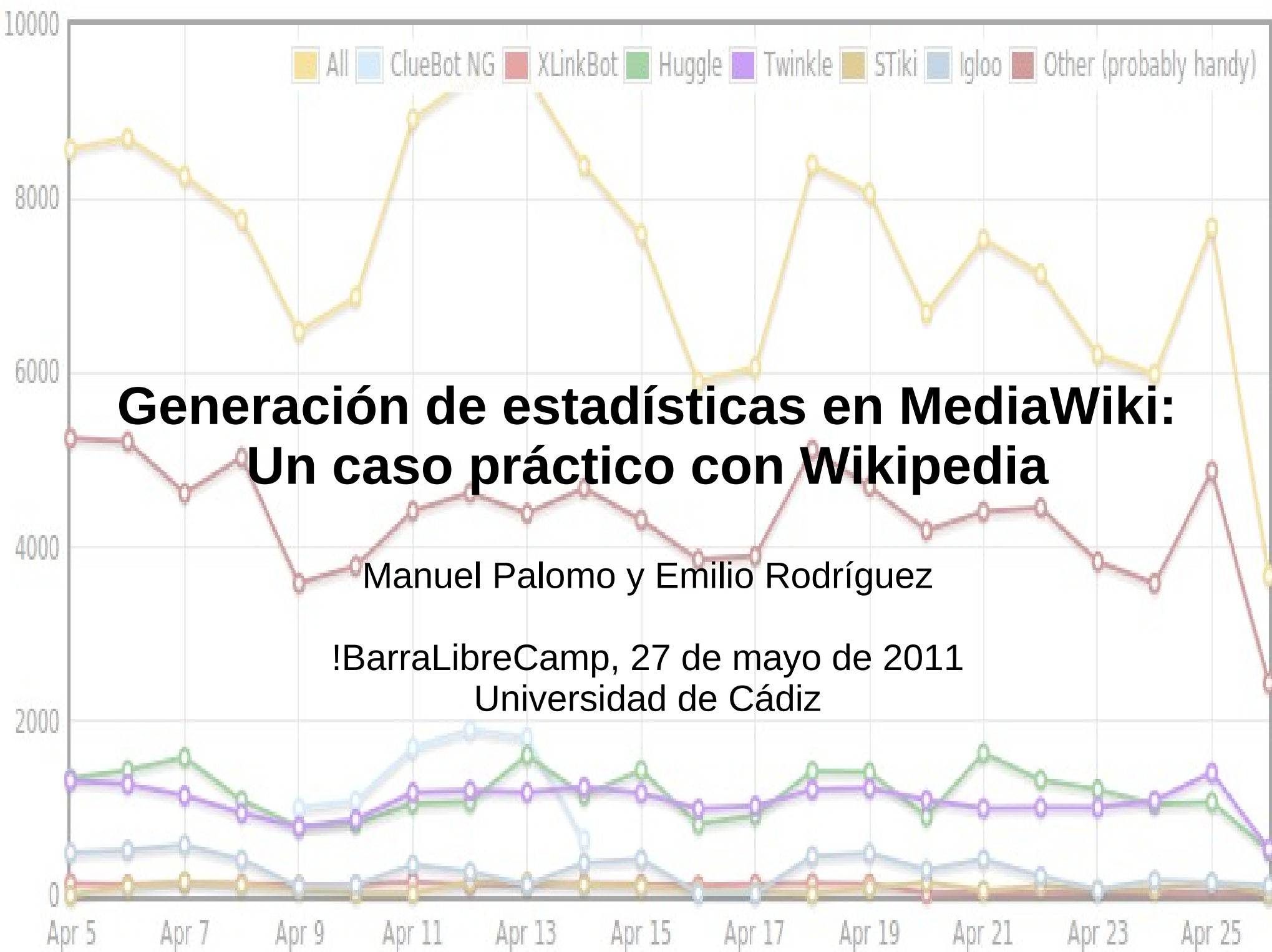


Generación de estadísticas en MediaWiki: Un caso práctico con Wikipedia

Manuel Palomo y Emilio Rodríguez

!BarraLibreCamp, 27 de mayo de 2011
Universidad de Cádiz



¿Qué es MediaWiki?

- Motor wiki avanzado y estable (inicio 2002)
 - Permite modificar las páginas del sitio web
- PHP+MySQL, libre, traducido a 300 idiomas
- Lo usan la mayoría de wikis de Internet
- Es ampliable con extensiones (hay cientos)
- Ejemplos:
 - Wikipedia (<http://wikipedia.org>)
 - Wikia (<http://www.wikia.com>)
 - Wikis de la OSLUCA (<http://osl.uca.es/wikis>)

¿Por qué importan las estadísticas?

- Ayudan a comprender cómo funciona un wiki
- Vemos dónde destacamos y dónde no
- Wikis asignaturas: ayudan a evaluar alumnos
 - WikiHaskell, WikiRA, WikiCE
- Ayudan a buscar soluciones y mejorar en el futuro
 - ¿Usuarios se van? ¿Por qué?
- *For fun* (por diversión)

Tipos de estadísticas

- Nativas de MediaWiki
 - Special:Statistics
 - Son bastante limitadas
- Generadas por el usuario
 - Usando datos crudos extraídos del wiki
 - Usando herramientas sobre estos datos:
 - Herramientas genéricas
 - Herramientas específicas
 - Llegan hasta donde tú quieras llegar

Estadísticas nativas

- Special:Statistics
- Número de:
 - Páginas
 - Usuarios
 - Imágenes
 - Ediciones



The screenshot shows the Wikipedia Special:Statistics page. On the left is a sidebar with the Wikipedia logo and navigation links. The main content area is titled 'Statistics' and includes a table with page and edit statistics, and another table with user statistics.

Special page

Statistics

From Wikipedia, the free encyclopedia

Page statistics	
Content pages	3,621,541
Pages (All pages in the wiki, including talk pages, redirects, etc.)	23,813,957
Uploaded files	844,800

Edit statistics	
Page edits since Wikipedia was set up	458,312,380
Average edits per page	19.25

User statistics	
Registered users	14,433,451
Active registered users (list of members) (Users who have performed an action in the last 30 days)	149,068
Bots (list of members)	679
Administrators (list of members)	1,786

Left Sidebar:

- WIKIPEDIA The Free Encyclopedia
- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia
- Interaction
 - Help
 - About Wikipedia
 - Community portal
 - Recent changes
 - Contact Wikipedia
- Toolbox
 - Upload file
 - Special pages

- Cosas sencillas como número de enlaces, ediciones por categorías, no las da

<http://en.wikipedia.org/wiki/Special:Statistics>

Estadísticas nativas

- Estadísticas nativas aplicadas a Wikipedia
 - +18 millones de artículos
 - +250 idiomas
 - +1000 millones de ediciones en 10 años
 - +28 millones de usuarios registrados
 - +10 millones de imágenes libres en Commons
 - +200 mil ediciones al día
 - imagina un proyecto con esos commits al día

Estadísticas generadas por usuario

- 1) Obtener los *datasets*
 - Dumps XML (<http://dumps.wikimedia.org>)
 - API (<http://en.wikipedia.org/w/api.php>)
 - Base de datos en vivo (*shell access*)
- 2) Herramientas genéricas
 - R (te haces tus propios *scripts*)
 - Python + Matplotlib + NumPy
- 3) Herramientas específicas
 - HistoryFlow, WikiXRay, StatMediaWiki, EvalMediaWiki

HistoryFlow

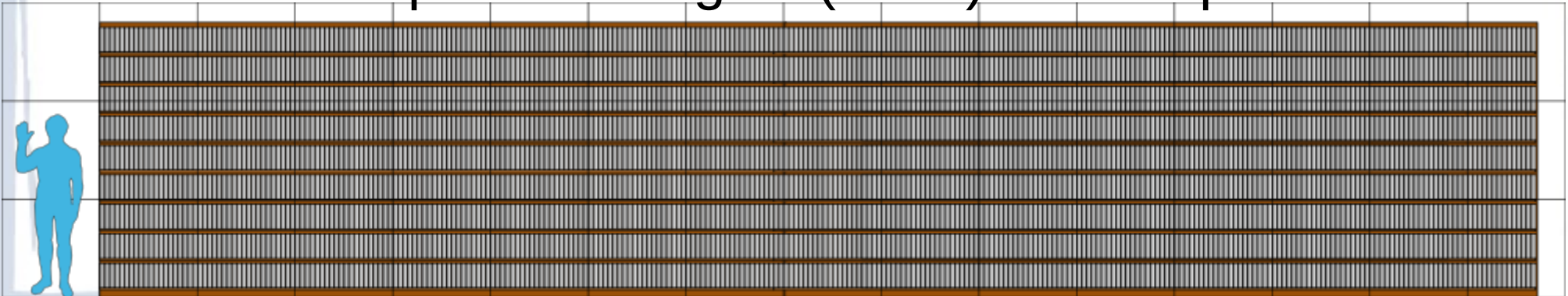
WikiXRay

StatMediaWiki

EvalMediaWiki

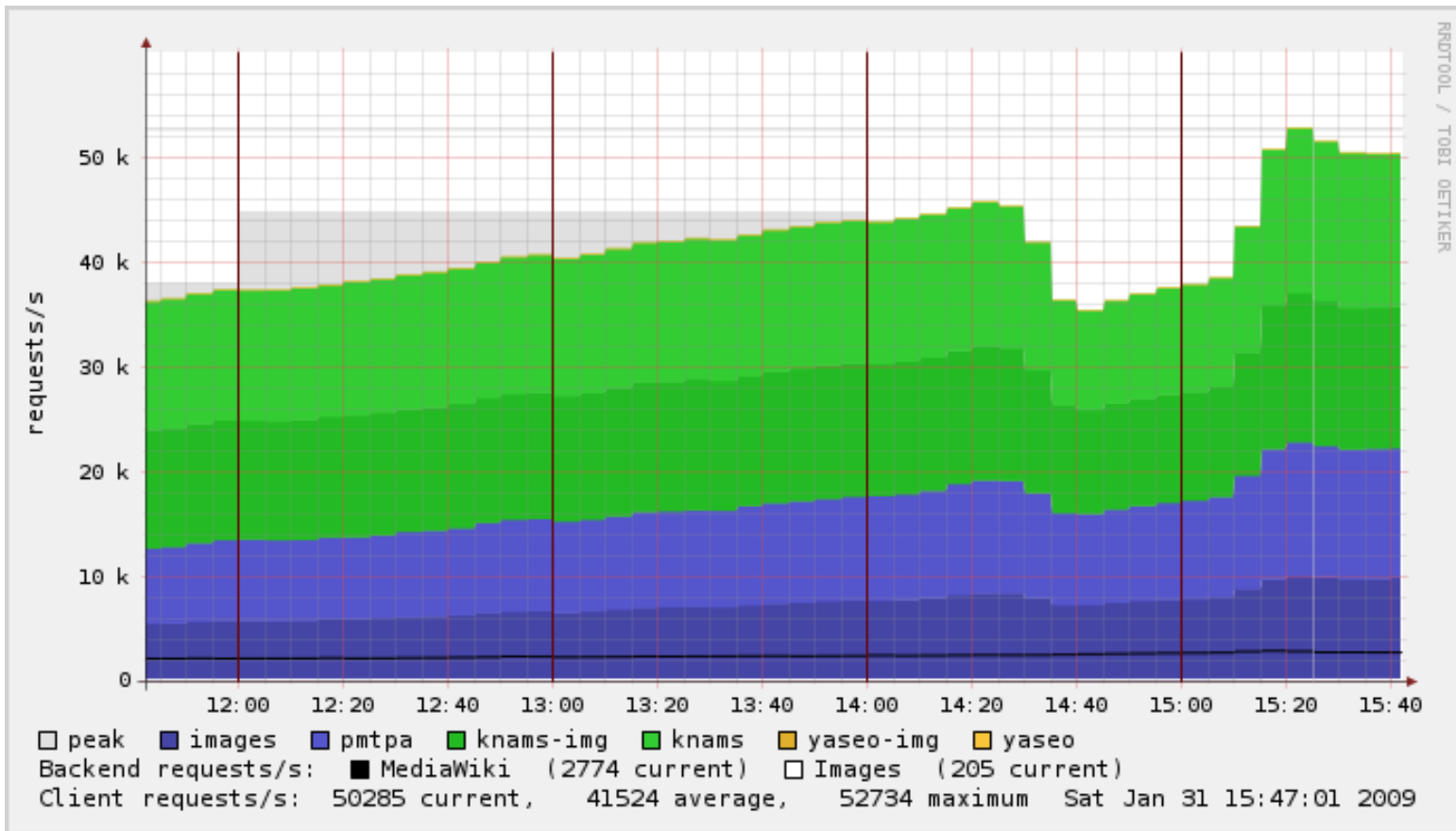
For fun

- No todo son estadísticas para investigación. Algunos ejemplos divertidos:
 - Contador en vivo de ediciones a Wikipedia
 - <http://toolserver.org/~emijrp/wikimediacounter/>
 - Geolocalización de editores anónimos
 - <http://www.lkozma.net/wpv/index.html>
 - Wikipedia en inglés (3.5M) si se imprimiese



For fun

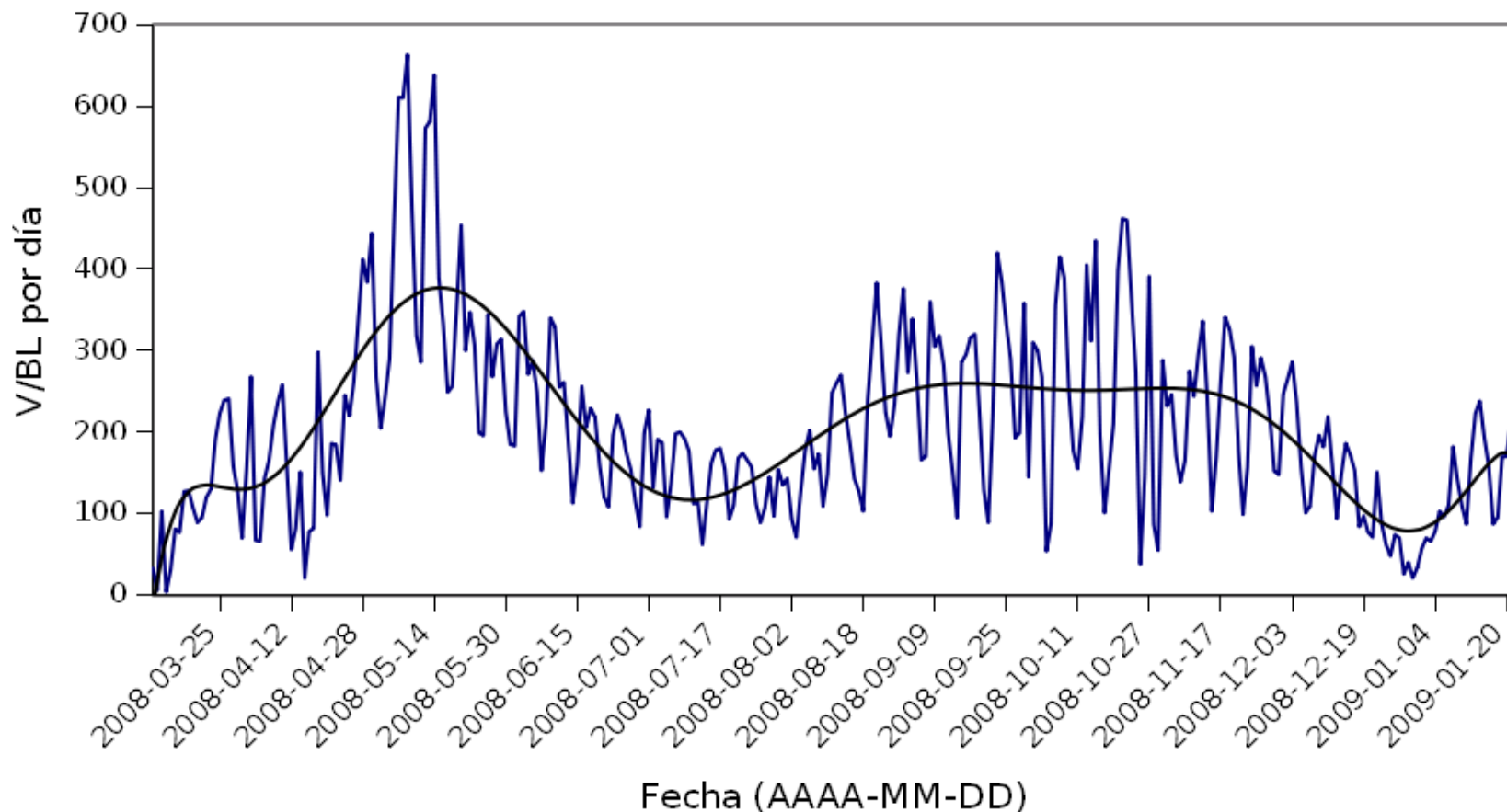
- Error de Google deja sin visitas a Wikipedia



<http://emijrp.blogspot.com/2009/01/el-25-de-los-hits-vienen-de-google.html>

For fun

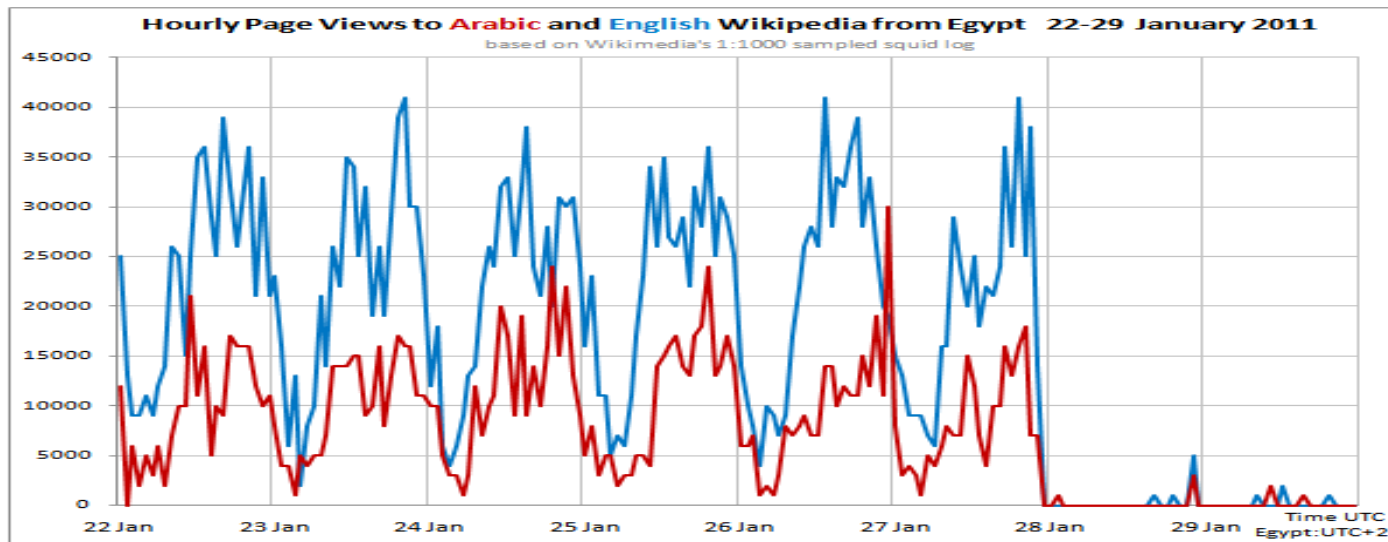
- En verano y fines de semana, vandalismo cae



<http://emijrp.blogspot.com/2009/04/la-sociologia-del-vandalismo.html>

For fun

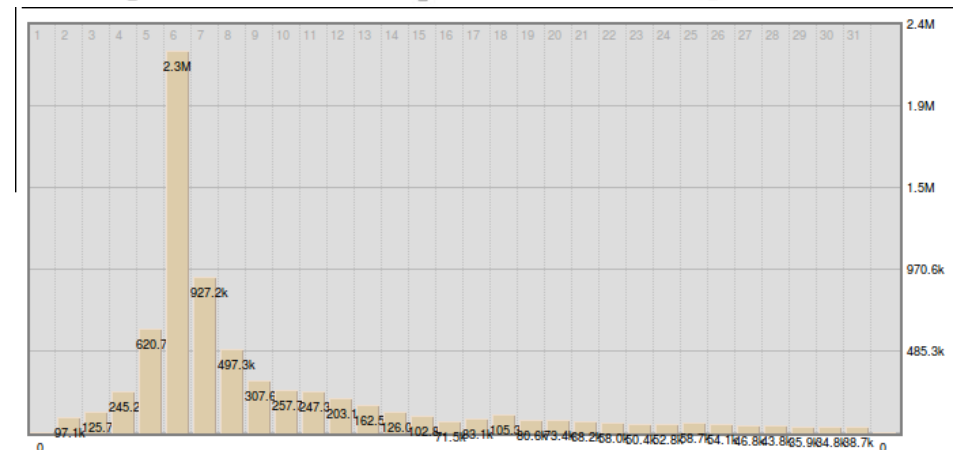
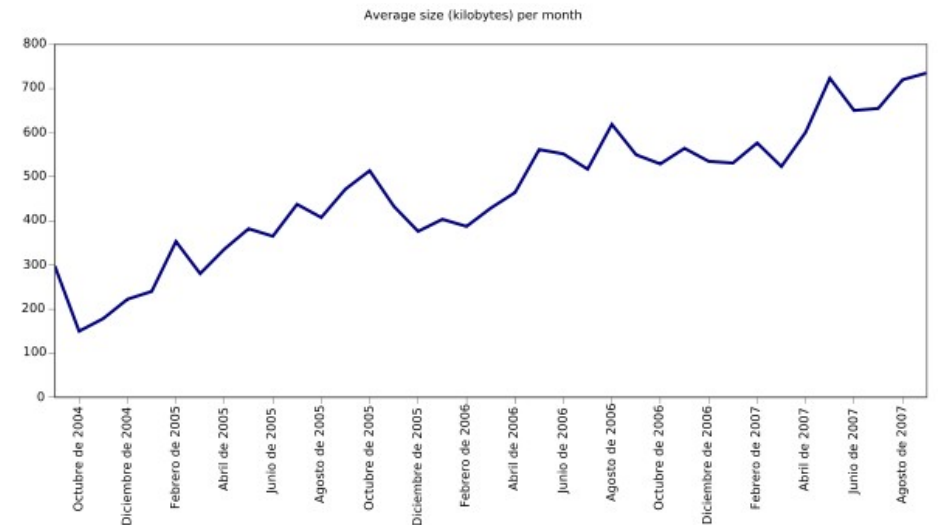
- Cada vez que Google cambia el logo, hay que proteger el artículo relacionado porque lo vandalizan en masa
- Corte de Internet en Egipto = ~ 0 visitas



<http://infodisiac.com/blog/2011/01/egypt-internet-switched-off/>

For fun

- Mejoran las cámaras...
...imágenes más KB
- Evento mundial...
...visitas en masa...
...y vandalismo



http://stats.grok.se/en/200811/Barack_Obama

<http://emijrp.blogspot.com/2007/10/ms-imgenes-y-mayor-resolucin.html>

Recursos sobre wiki estadísticas

- <http://en.wikipedia.org/wiki/Wikipedia:Statistics>
- <http://stats.wikimedia.org>
- <http://infodisiac.com/blog/>
- <http://stats.grok.se/>

Gracias por la atención

Join us on [Facebook](#) and follow us on [Twitter](#)!
Chat with us at [Freenode](#)

[Translate it!](#)

Total edits in [Wikimedia projects](#):

1,000,000,000

Press F11 for fullscreen

[Donate to Wikimedia Foundation](#)

Developed by [emjrp](#) (Inspired by [Zis7](#))

¿Preguntas?