

Análisis de patrones en la evolución de wikis.

Martin Jonathan y Torres Diego.

Cita:

Martin Jonathan y Torres Diego (2016). *Análisis de patrones en la evolución de wikis*. *Humanidades Digitales: Construcciones locales en contextos globales*. Asociación Argentina de Humanidades Digitales, Buenos Aires.

Dirección estable: <https://www.aacademica.org/aahd2016/19>

ARK: <https://n2t.net/ark:/13683/ey3x/SRz>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.



ASOCIACIÓN ARGENTINA DE HUMANIDADES DIGITALES

Humanidades Digitales: construcciones locales en contextos globales

Actas del I Congreso Internacional
de la Asociación Argentina de
Humanidades Digitales (AAHD)



Humanidades Digitales : Construcciones locales en contextos globales : Actas del I Congreso Internacional de la Asociación Argentina de Humanidades Digitales - AAHD / Agustín Berti ... [et al.] ; editado por Gimena del Rio Riande, Gabriel Calarco, Gabriela Striker y Romina De León - 1a ed. - Ciudad Autónoma de Buenos Aires : Editorial de la Facultad de Filosofía y Letras Universidad de Buenos Aires, 2018.

Libro digital, PDF

Archivo Digital: descarga y online

ISBN 978-987-4019-97-4

1. Actas de Congresos. 2. Humanidades. 3. Digitalización. I. Berti, Agustín II. del Rio Riande, Gimena, ed.

CDD 301

ISBN 978-987-4019-97-4



Humanidades Digitales. Construcciones locales en contextos globales

Gimena del Rio Riande, Gabriel Calarco, Gabriela Striker y Romina De León (Eds.)

ISBN: 978-987-4019-97-4

> Índice

I. Preliminares

FUNES, Leonardo. *Palabras Preliminares*

del **RIO RIANDE**, Gimena. *Cuando lo local es global*

FIORMONTE, Domenico. *¿Por qué las Humanidades Digitales necesitan al Sur?*

II. Métodos y herramientas de las Humanidades Digitales

BIA, Alejandro. *Estilometría computacional, algunas experiencias en el marco del proyecto TRACE*

SALERNO, Melisa; **HEREÑÚ**, Daniel y **RIGONE**, Romina. *Modelado 3D del cementerio de la antigua Misión Salesiana de Río Grande: tareas efectuadas y potenciales usos*

VÁZQUEZ CRUZ, Adam Alberto y **TAYLOR**, Tristan. *Adnoto: un etiquetador de textos para facilitar la creación de ediciones digitales*

BRACCO, Christian; **CORREA**, Facundo; **CUEVAS**, Lucas; **CEPEDA**, Virginia; **DELLEDONNE**, Francisco; **VOSKUIL**, Anne Karin; **PAPARAZZO**, Nicolás y **TORRES**, Diego. *Una wiki semántica para las artes escénicas. Conceptos e implementación de la plataforma colaborativa Nodos*

IZETA, Andrés Darío y **CATTÁNEO**, Roxana. *¿Es posible una arqueología digital en Argentina? Un acercamiento desde la práctica*

LACALLE, Juan Manuel y **VILAR**, Mariano. *Una lectura distante de la investigación actual en*

Letras en Argentina

MARTIN, Jonathan y **TORRES**, Diego. *Análisis de patrones en la evolución de wikis*

MARTÍNEZ CANTÓN, Clara Isabel; **DEL RIO RIANDE**, Gimena y **GONZÁLEZ-BLANCO GARCÍA**, Elena. *Poetriae. Una colección de poéticas medievales basada en conceptos métricos únicos y referenciables*

SUED, Gabriela. *Ciudades visibles: estética y temática de tres ciudades iberoamericanas en la red social Instagram. Un estudio exploratorio desde las Humanidades Digitales*

III. Educación, políticas públicas, Humanidades Digitales en el aula

DAVICO, María Luz; **LINEARES**, Gabriel y **PEZZUTTI**, Luciana. *Literacidad electrónica en la enseñanza universitaria: cómo, cuándo y dónde*

MUÑOZ, Patricia Alejandra. *Valoración de un proyecto de desarrollo tecnológico y social en la enseñanza de Inglés como lengua extranjera*

PACHECO DE OLIVEIRA, Maria Livia y **SÁ DE PINHO NETO**, Júlio Afonso. *Brecha digital e o acesso à informação: projetos de inclusão digital*

CASASOLA, Laura. *Experiencia educativa con TIC: Celulares en acción*

DÍAZ, Aída Alejandra y **HUALPA**, Mariela. *Una experiencia de aprendizaje en educación superior mediada por TIC*

FRESCURA TOLOZA, Claudio Daniel. *Computación en la nube en la enseñanza de escritura académica*

LEÁNEZ, Nancy; **LECETA**, Andrea; **MARTÍN**, Marcela y **MORCHIO**, Marcela. *Hacia una reconfiguración del aula de lengua extranjera*

OLAIZOLA, Andrés. *Los escritores vernáculos digitales y el concepto de valor en las escrituras digitales*

CHECHELE, Patricia; **LURO**, Vanesa y **PINTOS ANDRADE**, Esteban. *Afiliarse en la distancia. El ingreso a la educación superior en un entorno virtual de aprendizaje*

ALLÉS TORRENT, Susanna y **DEL RIO RIANDE**, Gimena. *Enseñar edición digital con TEI en español. Aprendizaje situado y transculturación*

IV. Medios, re-mediación, redes sociales

RODRIGUEZ KEDIKIAN, Martín. #100DiasdeMacri. *Analítica cultural en la construcción de los primeros cien días de la presidencia de Mauricio Macri en conversaciones en Twitter*

ALONSO, Julio; **ALAMO**, Sofía; **GONZALEZ OCAMPO**, María Eugenia; **GIAMBARTOLOMEI**, Guido; **MANCHINI**, Lucas y **TOSCANO**, Ayelén. *¿Hacia una algoritmización de los sentimientos?*

DE MIRANDA, Jair Martins. *Samba Global– Do mundo do samba ao samba no mundo*

ORTIZ, María. *Las migraciones en los tiempos del software*

SANTOS, Laura. *Arte urbano, de la calle a las redes*

ALAMO, Sofía; **BORDOY**, Giselle; **CHETTO**, Melisa; **IBAÑEZ**, Fernanda, **MIGLIORINI**, Agustina y **GONZALEZ OCAMPO**, María Eugenia. #NiUnaMenos: *Big Data para la comprensión de una problemática de género*

KLIMOVSKY, Pedro. *El documental digital y la representación de lo real*

BERTI, Agustín. *Fotogramas autorizados: La crisis de la noción de obra cinematográfica ante las remasterizaciones*

BORDOY, Giselle. *El disco como obra abierta en interacción con las audiencias*

COELHO, Cidarley. *Forma Material Digital: livro e leitura na sociedade contemporânea*

V. Reflexiones sobre/desde/hacia lo digital

VISCARDI, Ricardo. *Actuvirtualidad e inter-rogação: un lugar entre-otros*

ÁLVAREZ GANDOLFI, Federico y **DEL VIGO**, Gerardo Ariel. *Hatsune Miku, una idol digital: entre el otakismo y el waifuismo*

SAÁ, Guido. *Reflexiones sobre música y narración: Recursos retóricos y exegéticos musicales en la línea narrativa y el pathos en BioShock 2 y BioShock Infinite*

GLUZMAN, Georgina Gabriela. *Algunas reflexiones sobre la Base de datos de mujeres artistas en Buenos Aires (1924-1939)*

DOMINGUEZ HALPERN, Estela; **ALAMO**, Sofía; **ALONSO**, Julio. *Entramados y ciudades. Visibilizando Baldosas por la Memoria*

GÓMEZ, Verónica Paula. *Territorios nacionales, territorialidades ciberespaciales: disputas discursivas sobre la soberanía en la circulación de literatura digital*

RIGAL COLLADO, Pablo Alonso; **MAESTIGUE PRIETO**, Nancy y **GARCÍA VÁZQUEZ**, Mayté. *La narración hipertextual. El reto cubano*

VI. La publicación científica y el Acceso Abierto desde las Humanidades Digitales

TSUJI, Teresa y **CANELLA**, Rubén. *Lenguajes y recursos multimediales para la difusión de la ciencia. Desafíos y oportunidades digitales*

CATALDI, Marcela; **DI CÉSARE**, Victoria; **FERNÁNDEZ**, Néstor; **HERNÁNDEZ**, Alicia; **LIBERATORE**, Gustavo y **VOUTTO**, Andrés. *Sistema taxonómico de organización de los recursos de información autoarchivados en el Repositorio Institucional de la Facultad de Humanidades de la Universidad Nacional de Mar del Plata*

ÁLVAREZ, Leonardo Javier y **CORDA**, María Cecilia. *FLACSOAndes Tesis: comunicación científica de investigaciones realizadas en maestrías y doctorados del sistema FLACSO*

VII. Digitalización, políticas y prácticas, archivo y memoria

AUTHIER, Carlos; **GIORDANINO**, Eduardo y **LUIRETTE**, Carlos. *La preservación de la memoria audiovisual en Argentina*

GAMBA, Guido; **HEIDEL**, Evelin; **RAIA**, Matías; **ACUÑA**, Ezequiel; **ACTIS CAPORALE**, Carla; **DE LA HERA**, Diego y **ACEVEDO**, Melisa. *Recursos digitales para el acceso a los bienes culturales en dominio público*

FLORES MUTIGLIENGO, Jennifer. *Arte y Archivo*

BUGNONE, Ana y **SANTAMARÍA**, Mariana. *La política de democratización del archivo: el caso del Centro de Arte Experimental Vigo*

GAMBA, Guido; **HEIDEL**, Evelin; **RAIA**, Matías; **ACUÑA**, Ezequiel; **ACTIS CAPORALE**, Carla; **DE LA HERA**, Diego y **ACEVEDO**, Melisa. *Digitalización: Una experiencia de campo*

Análisis de patrones en la evolución de wikis

MARTIN, Jonathan / Laboratorio de Investigación y Formación en Informática Avanzada.
Universidad Nacional de La Plata (LIFIA, UNLP) – Jonathan.Martin@lifa.info.unlp.edu.ar

TORRES, Diego / Laboratorio de Investigación y Formación en Informática Avanzada.
Universidad Nacional de La Plata (LIFIA, UNLP). Departamento de Ciencia y Tecnología.
Universidad Nacional de Quilmes (UNQ) – diego.torres@lifa.info.unlp.edu.ar

» Palabras clave: evolución de wikis, análisis estructural, patrones, visualización, Wikimedia.

» Resumen

En este trabajo se plantea un estudio sobre la evolución en la generación de conocimiento en el contexto de wikis. Una wiki es un sitio web de edición colaborativa donde cualquier usuario puede editar, crear o modificar contenido directamente usando su navegador. Al ser netamente colaborativos, las wikis reúnen la participación conjunta de muchas personas, en algunos casos miles o millones, que comparten un mismo objetivo. Es por ello y por ser plataformas de fácil acceso, orientadas a la generación de conocimiento y de gran actividad, que resultan un campo de gran interés. Además, la gran cantidad de personas que participan colaborativamente, por ejemplo, en las wikis nucleadas por *Wikimedia*, hacen que su empleo sea un fenómeno en el que confluyen personas con diferentes formaciones, ubicaciones geográficas y culturas. Por otro lado, las wikis poseen un *registro de revisiones* en el cual se guardan el momento y el autor de cada uno de los cambios realizados como pueden ser las ediciones del contenido de las páginas, la creación de estas y de sus categorías, entre otras modificaciones. En función de lo anterior se presenta un enfoque para poder analizar la evolución en el contenido de una wiki a través de su registro de revisiones. Este estudio nos permite detectar y clasificar patrones sobre la evolución en el contenido de las wikis. Se plantea analizar ciertos cambios estructurales como la creación de páginas o nuevos links, como así también la identificación de cambios e interpretaciones en el contenido de los artículos. Finalmente, generar diferentes visualizaciones que muestran la evolución a partir de gráficos e información estadística. Este trabajo se centra en el estudio aplicado a la tecnología *MediaWiki*, motor de las wikis más importantes de la actualidad como *Wikipedia*.

> **Introducción**

Una wiki es un sitio web de edición colaborativa que permite a los usuarios crear y editar contenido de forma dinámica. La principal característica de las wikis es la flexibilidad con la cual se puede generar este contenido de forma simple a partir de un editor de texto que brinda la wiki y que utiliza un sistema de *markup* o marcado de texto para dar formato al contenido o estructurarlo.

La primera wiki fue creada por el programador Ward Cunningham. Su intención al crear este software, al que llamó *WikiWikiWeb*, fue mejorar la generación de documentación de software a partir de una edición sencilla que pudiera realizarse desde un editor de texto que permitiera la creación y modificación de contenido de forma colaborativa y rápida (Cunningham *et al.*, 2014).

Dado que se pueden registrar casos en los que se generen ediciones erróneas o las consideradas ediciones vandálicas, que son las ediciones cuya intención es maliciosa, las wikis poseen por cada artículo un historial de todos los cambios que se realizaron.

Este historial de cambios es conocido como *historial de revisiones (revision history)* (véase Figura 1). El usuario puede acceder a este a partir de la opción *View History*, dentro puede observar un listado con las revisiones que se realizaron de la más reciente a la más antigua. Esta historia posee información sobre cuándo se realizó, quién la realizó y sobre su contenido final. También nos permite seleccionar dos revisiones para comparar el contenido de ambas a fin de registrar los cambios entre dos revisiones.

Este trabajo tiene como objetivo visualizar la evolución de los artículos de una wiki, a partir de lo cual se busca detectar y clasificar patrones sobre la evolución en el contenido de la wiki. Se plantea analizar ciertos cambios estructurales como la creación de páginas o nuevos links, como así también la identificación de cambios e interpretaciones. Finalmente, generar diferentes visualizaciones que muestran la evolución a partir de gráficos e información estadística.



Nupedia: Revision history

Help

View logs for this page

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia
- Wikipedia store

- Interaction
- Help
- About Wikipedia
- Community portal
- Recent changes
- Contact page

- Tools
- What links here
- Related changes
- Atom
- Upload file
- Special pages
- Page information
- Wikidata item

Languages

Browse history

From year (and earlier): 2016 From month (and earlier): all Tag

filter: Show

For any version listed below, click on its date to view it.

For more help, see [Help:Page history](#) and [Help:Edit summary](#).

External tools: [Revision history statistics](#) · [Revision history search](#) · [Edits by user](#) · [Number of watchers](#) · [Page view statistics](#)

(cur) = difference from current version, (prev) = difference from preceding version, m = minor edit, → = section edit, ← = automatic edit summary

(newest | oldest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

Compare selected revisions

- [\(cur | prev\)](#) 03:38, 18 November 2016 GreenC bot (talk | contribs) **m** . . (12,668 bytes) (+49) . . (1 archive template merged to [{{webarchive}}](#) (WAM)) (undo)
- [\(cur | prev\)](#) 10:57, 8 November 2016 NasssaNser (talk | contribs) **m** . . (12,619 bytes) (-21) . . (undo)
- [\(cur | prev\)](#) 04:59, 29 October 2016 68.194.91.23 (talk) . . (12,640 bytes) (-4) . . (→History) (undo)
- [\(cur | prev\)](#) 13:44, 26 October 2016 Znrdrig (talk | contribs) . . (12,644 bytes) (+1) . . (fixed spelling of Ruth lfcher from Ruth Ifner) (undo) (Tag: Visual edit)
- [\(cur | prev\)](#) 18:01, 25 October 2016 Dawniraci (talk | contribs) . . (12,643 bytes) (-11) . . (→Editorial process: fixed a bit more to the Ruth Ifner paragraph) (undo)
- [\(cur | prev\)](#) 17:53, 25 October 2016 Dawniraci (talk | contribs) . . (12,654 bytes) (+421) . . (→Editorial process: added information from my textbook about Ruth Ifner) (undo)
- [\(cur | prev\)](#) 13:42, 14 September 2016 GreenC bot (talk | contribs) **m** . . (12,233 bytes) (+1) . . (WaybackMedic 2) (undo)
- [\(cur | prev\)](#) 21:36, 12 September 2016 Zigzig20s (talk | contribs) . . (12,232 bytes) (+535) . . (added more referenced info with a direct quote from a book I'm reading) (undo)
- [\(cur | prev\)](#) 11:49, 18 August 2016 Bender the Bot (talk | contribs) **m** . . (11,697 bytes) (+2) . . (→Further reading: http→https for Internet Archive (see this RfC) using AWB) (undo)
- [\(cur | prev\)](#) 05:14, 18 August 2016 MB298 (talk | contribs) . . (11,695 bytes) (+36) . . (→Editorial process: fix) (undo)

Figura 1. Historial de revisiones del artículo Nupedia. Fuente: <https://en.wikipedia.org/>.

> Desarrollos propuestos

Se desarrollará un software que permita en primera instancia la obtención de la información histórica de una wiki a partir de los *dumps*, o copias de respaldo, en XML generados mensualmente. Con esa información se propondrá un modelo, basado en el paradigma orientado a objetos, que permita el desarrollo de estrategias para:

- Presentar información de los cambios entre diferentes revisiones de los artículos.

- Visualizar de forma gráfica y textual la información sobre el estado al momento del dump de la wiki respecto a su contenido y su distribución según aspectos como autoría o estructurales como pertenencia a sus respectivos namespaces y categorías.

Informar sobre la evolución del contenido de la wiki en el tiempo.

› **Definiciones preliminares y métricas**

A partir de la información obtenida se desea evaluar las mismas con un conjunto de métricas, las cuales se utilizarán posteriormente para la presentación gráfica de la información.

Comenzaremos definiendo una wiki (W), perteneciente a la familia de las *MediaWikis*, mediante la fórmula $W = (N; P)$ en la que una wiki contiene un conjunto de *namespaces* (N) y un conjunto de páginas (P).

Cada página de P contiene información del *namespace* al que pertenece, el cual nunca cambia una vez creado, y posee una colección de revisiones que le fueron realizadas de forma tal que definimos una página (p) usando la fórmula $p = (n; R)$, teniendo en cuenta que $p \in P$ y siendo R el conjunto de revisiones de la página y n el *namespace* de la página siempre que $n \in N$.

Cada revisión de R contiene quién la realizó y cuándo, el texto final de la revisión y, si existe una revisión anterior, también hace referencia a esta. Por lo cual, y a partir de considerar que $r \in R$, podemos definir la revisión como $r = (id; pr; d; a; t)$ siendo id un identificador, que es un número natural mayor a 0 y único para su identificación; pr la revisión anterior, o padre, identificada por un número natural mayor a 0 o negativo si r no posee revisión anterior; además, siendo d la fecha en que se realizó la revisión; a el autor de la misma y t el texto de la página para esta revisión.

Para referirnos al conjunto de revisiones de la página p utilizaremos R_p y para una revisión particular de la página p utilizaremos r_p siempre que $r_p \in R_p$. También para una revisión r señalaremos su revisión anterior, autor, fecha en que se realizó y texto como pr_r, a_r, d_r y t_r , respectivamente.

Además, en este trabajo para referirnos al cardinal de un conjunto B se utilizará la notación $card(B)$. A partir de esto se definieron las siguientes métricas:

#RevisionsOfAuthor(a, R):

Esta métrica indica la cantidad de revisiones realizadas por el autor a en la colección de revisiones R .

$card(\{r / r \in R \wedge a_r = a\})$

Con esta métrica podemos saber, por ejemplo, cuántas revisiones realizaron exactamente cada uno de los autores a_1, a_2 y a_3 de un conjunto de revisiones.

PercentOfAuthorRevisions(a, R):

Esta métrica representa el porcentaje de revisiones realizadas por el autor a de la colección R de revisiones.

$$\frac{\#RevisionsOfAuthor(a, R) * 100}{R}$$

Con esta métrica, dado un conjunto de revisiones, podemos saber cuál fue el porcentaje de las realizadas por un autor determinado.

PercentOfRevisionByAuthor(p):

A partir de las métricas mencionadas anteriormente se puede calcular cuál fue el aporte de cada autor a una página p con la siguiente métrica.

$$\{(a_r, PercentOfAuthorRevisions(a_r, R_p)) / r \in R_p\}$$

Con esta métrica para un determinado artículo podemos conocer por cada autor que realizó ediciones qué porcentaje de estas hizo sobre el total que posee el artículo.

RevisionsOfDay(d, R):

Con esto podemos obtener las revisiones pertenecientes a la colección R que hayan sido realizadas en la fecha d .

$$\{r / r \in R \wedge d_r = d\}$$

RevisionPerDay(p):

Esta métrica nos permite obtener las revisiones que se realizaron para cada día en el que al menos se realizó una revisión. Además, a partir de esto podemos saber la cantidad de las que se realizaron para un día determinado.

$$\{(d_r, RevisionsOfDay(d_r, R_p)) / r \in R_p\}$$

BytesOfRevisionsPerDay(p):

Asumiendo que $\#t$ nos da el tamaño del texto en bytes, esto permite obtener por día los cambios de longitud en el texto de una página p .

$$\{(d_r, \#t_r) / r \in R_p\}$$

Además de los componentes propios de las páginas que obtenemos de las wikis como las revisiones o *namespaces*, también extraemos de los textos de las revisiones una colección de estilos, nombrada S , tal que S_t está formada por un conjunto de *tuplas* que contienen un estilo s y las ocurrencias de dicho estilo en t .

$$S_t = \{(s, T) / s \in S \wedge T = \{t_i / t_i \in t\}\}$$

Por ejemplo, a partir de un fragmento de texto como el extraído del artículo “Clasificación TAS” (véase Figura 2) podemos decir que el conjunto S_t contendría solo dos elementos: uno que representa a los encabezados de nivel 2 y, el segundo, a los enlaces internos, y cada uno con su respectivo conjunto de ocurrencias.

== Uso de la clasificación TAS ==

Debe observarse, como se discute con detalle por Le Maitre y otros (2002), que esta clasificación no puede aplicarse a todas las [[roca volcánica|rocas volcánicas]]. Ciertas rocas no pueden nombrarse usando el [[diagrama]]. Para otras, se deben usar criterios adicionales mineralógicos, químicos, y de textura, como por ejemplo con los [[lamprófiro]]s.

Figura 2. Extracto del artículo "Clasificación TAS" de Wikipedia. Fuente: https://es.wikipedia.org/wiki/Clasificaci%C3%B3n_TAS.

#OcurrencesOfStyle(s, t):

Esta métrica permite obtener la cantidad de ocurrencias del estilo s en el texto t .

$$\text{card}(\{i_r / i_s \in S_t \wedge i_s = s\})$$

De esta forma podemos obtener métricas basadas en esta.

StylesInText(t):

Con esta métrica obtenemos un listado con la cantidad de ocurrencias de cada estilo en el texto t .

$$\{(s, \#OcurrencesOfStyle(s, t)) / s \in S_t\}$$

StylesOfRevisionsPerDay(p):

Siendo $\#S_t$ la cantidad de elementos de la colección S_t obtenemos por revisión la cantidad de estilos que posee y la fecha de la revisión.

$$\{(d_r, \sum_i^{StylesInText(t_r)} i_T) / r \in R_p\}$$

› Fuentes de información

Dado que toda la información que es de nuestro interés para analizar se encuentra dentro de las revisiones, nuestro primer enfoque es cómo obtener las mismas para nuestro análisis.

En este trabajo se encontraron dos fuentes de información que nos aportaban datos sobre el total de las revisiones para los artículos tanto de forma individual utilizando la API

Special:Export que brindan las wikis como de modo integral a través del procesamiento de *dumps* de la *Wiki*, lo cual nos permite obtener todas las revisiones del conjunto de los artículos de una wiki.

Respaldos de MediaWikis

El motor de wikis *MediaWiki* brinda la posibilidad a las wikis de exportar periódicamente todo su contenido hacia la forma de respaldo. Estos respaldos conocidos como *dumps* quedan accesibles desde la web de *Wikimedia* (véase Figura 3).

Wikimedia Downloads

Please note that we have rate limited downloaders and we are capping the number of per-ip connections to 2. This will help to ensure that everyone can access the files with reasonable download times. Clients that try to evade these limits may be blocked.

Data downloads

The Wikimedia Foundation is requesting help to ensure that as many copies as possible are available of all Wikimedia database dumps. Please **volunteer to host a mirror** if you have access to sufficient storage and bandwidth.

Database backup dumps
A complete copy of all Wikimedia wikis, in the form of wikitext source and metadata embedded in XML. A number of raw database tables in SQL form are also available.
These snapshots are provided at the very least monthly and usually twice a month. If you are a regular user of these dumps, please consider subscribing to [xmldata.dumps](#) for regular updates.

Mirror Sites of the XML dumps provided above
Check the [complete list](#).

Static HTML dumps
A copy of all pages from all Wikipedia wikis, in HTML form.
These are currently not running.

DVD distributions
Available for some Wikipedia editions.

Backup dumps of wikis which no longer exist
A complete copy of selected Wikimedia wikis which no longer exist and so which are no longer available via the main database backup dump page. This includes, in particular, the Sept. 11 wiki.

Analytics data files
Pageview, Mediaccount, Unique, and other stats.

Other files
Image tarballs, survey data and other items.

Kiwix files
Static dumps of wiki projects in OpenZim format

Dataset collection at the Data Hub (off-site)
Many additional datasets that may be of interest to researchers, users and developers can be found in this collection.

Software downloads

MediaWiki
MediaWiki is a free software wiki package written in PHP, originally for use on Wikipedia. It is now used by several other projects of the non-profit Wikimedia Foundation and by many other wikis.

Figura 3. Página principal de Wikimedia. Fuente: <https://dumps.wikimedia.org/>.

Desde esta web podemos acceder a los respaldos de todas las *MediaWikis* entrando a *Database backup dumps*, a partir de allí podemos seleccionar qué wiki es de nuestro interés para obtener sus *dumps*.

Una vez seleccionada la wiki nos encontraremos con que no hay un solo tipo de *dump*, sino que tenemos una gran variedad de archivos para descargar. A continuación, se describen algunos de estos (véase Tabla 1).

Archivo	Contenido
<i>pages-articles.xml</i>	Este archivo contiene todos los artículos solo con su revisión más reciente, también posee información sobre <i>templates</i> y descripción de archivos.
<i>pages-logging.xml</i>	Este archivo contiene información sobre creación o bloqueo de usuarios, cargas de imágenes, importación o movimiento de páginas, aumento de niveles de protección y eliminación de revisiones.
<i>pages-meta-current.xml</i>	Este archivo contiene información sobre páginas personales de usuarios y de discusión.
<i>pages-meta-history.xml</i>	Este archivo posee información de todos los artículos de todos los <i>namespaces</i> de la wiki con todas las revisiones que se le realizaron a cada uno.
<i>stub-articles.xml</i>	Este archivo contiene metadata de todos los artículos solo con su revisión más reciente, también posee información sobre <i>templates</i> y descripción de archivos. No posee texto en las revisiones, solo la longitud de las mismas.
<i>stub-meta-current.xml</i>	Este archivo contiene solo metadata de páginas personales de usuarios y de discusión. No posee texto en las revisiones, solo la longitud de las mismas.
<i>stub-meta-history.xml</i>	Este archivo posee metadata de todos los artículos de todos los <i>namespaces</i> de la wiki con todas las revisiones que se le realizaron a cada uno. No posee texto en las revisiones, solo la longitud de las mismas.

Tabla 1. Archivos de dumps. Fuente: <https://dumps.wikimedia.org/>.

Como se puede apreciar en la tabla 1, tenemos diversidad de archivos con gran variedad de información y distintos niveles de completitud. En nuestro caso nos interesa estudiar el archivo *pages-meta-history.xml* ya que posee el total de la información sobre los artículos y sus revisiones.

Una ventaja de esta forma de obtención de información es que contamos con el total de los artículos de la wiki para estudiar. Por otro lado, esto también es una desventaja porque dependiendo de qué wiki se tome para estudiar, el tamaño de estos archivos puede escalar rápidamente.

Funcionalidad de exportación en wikis

En las wikis que utilizan el motor *MediaWiki* se pueden obtener artículos en formato

XML a partir de la funcionalidad *Special:Export* de dos formas: una, a través de la interfaz web brindada por las wikis y, otra, mediante solicitudes POST.

A continuación, podemos observar la interfaz para exportar artículos de *Wikipedia* en inglés (véase Figura 4).

Special page Search Wikipedia

Export pages ? Help

You can export the text and editing history of a particular page or set of pages wrapped in XML. This may then be imported into another wiki running MediaWiki using *Special:Import*, if it is enabled. It may also be scanned by *AutoWikiBrowser*'s database scanner.

To export article pages, enter the titles in the text box below, one title per line, and select whether you want the current version as well as all old versions, with the page history lines, or just the current version with the info about the last edit. Other parameters of this interface, not available via this web form, are explained in *Parameters to Special:Export*.

To export the current revision of an article, a link such as *Special:Export/Train* can be used for the article *Train*.

Full history exports are limited to 1000 revisions.

Add pages from category:

Add pages manually:

Include only the current revision, not the full history

Include templates

Save as file

Figura 4. Menú para exportar páginas de Wikipedia. Fuente: <https://en.wikipedia.org/>.

En esta interfaz podemos solicitar artículos pertenecientes a una categoría o a partir del nombre del artículo, que es único y lo identifica, podemos listar los artículos que son de nuestro interés. Además, se nos permite optar por incluir solo la revisión actual o el historial de sus revisiones, o incluir su *templates* o no.

La posibilidad de realizar esta solicitud mediante una consulta POST también brinda las mismas opciones y, en adición, nos permite indicar a partir de qué fecha y hora queremos obtener las revisiones y la cantidad de las que queremos que acompañen al artículo.

En ambos casos existe un límite sobre el número que podemos obtener en una sola consulta, este límite es de 1000 revisiones por consulta. Mediante la interfaz web esta restricción siempre se aplicaría a obtener solo las 1000 más antiguas, mientras que en la solicitud POST esto dependería de la fecha y hora que le indiquemos como base.

Estructura de la información

Tanto si trabajamos con los *dumps* como con artículos exportados ambos se encuentran en formato XML, el cual es un formato de marcado que nos permite estructurar información. A continuación se muestra un extracto de un XML generado por la wiki. Cabe destacar que la estructura del XML es la misma sea obtenida por la funcionalidad de exportar o del *pages-meta-history.xml*.

```
<mediawiki xmlns="http://www.mediawiki.org/xml/export-0.10/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.mediawiki.org/xml/export-0.10/ http://www.mediawiki.org/xml/export-0.10.xsd" version="0.10" xml:lang="en">
  <siteinfo>
    <sitename>Wikipedia</sitename>
    <dbname>enwiki</dbname>
    <base>https://en.wikipedia.org/wiki/Main_Page</base>
    <generator>MediaWiki 1.29.0-wmf.4</generator>
    <case>first-letter</case>
    <namespaces>
      <namespace key="-2" case="first-letter">Media</namespace>
      <namespace key="-1" case="first-letter">Special</namespace>
      <namespace key="0" case="first-letter" />
      <namespace key="1" case="first-letter">Talk</namespace>
      <namespace key="2" case="first-letter">User</namespace>
      <namespace key="3" case="first-letter">User talk</namespace>
      <namespace key="4" case="first-letter">Wikipedia</namespace>
      <namespace key="5" case="first-letter">Wikipedia talk</namespace>
      <namespace key="6" case="first-letter">File</namespace>
      <namespace key="7" case="first-letter">File talk</namespace>
      <namespace key="8" case="first-letter">MediaWiki</namespace>
      <namespace key="9" case="first-letter">MediaWiki talk</namespace>
      <namespace key="10" case="first-letter">Template</namespace>
      <namespace key="11" case="first-letter">Template talk</namespace>
      <namespace key="12" case="first-letter">Help</namespace>
      <namespace key="13" case="first-letter">Help talk</namespace>
      <namespace key="14" case="first-letter">Category</namespace>
      <namespace key="15" case="first-letter">Category talk</namespace>
      <namespace key="100" case="first-letter">Portal</namespace>
      <namespace key="101" case="first-letter">Portal talk</namespace>
    ...
  </namespaces>
</siteinfo>
<page>
  <title>Pope</title>
  <ns>0</ns>
  <id>23056</id>
  <revision>
    <id>273292</id>
    <timestamp>2001-11-09T13:56:35Z</timestamp>
    <contributor>
      <username>Malcolm Farmer</username>
      <id>135</id>
    </contributor>
    <minor/>
    <comment>revert</comment>
    <model>wikitext</model>
  </revision>
</page>
</mediawiki>
```

```
<format>text/x-wiki</format>
<text xml:space="preserve" bytes="12105">
</text>
<sha1>0sbfeu2zdgzgtv0d7cxj2sv4o77707j</sha1>
</revision>
</page>
</mediawiki>
```

En el esquema anterior se presenta una única página con una única revisión a modo de ejemplo, también se omite el contenido de la revisión para facilitar la lectura. La estructura del XML está dada por un nodo raíz que representa a la wiki y está compuesto por $1+n$ nodos, es decir, un nodo que contiene información sobre la wiki como el nombre y los *namespaces* con sus identificadores y n nodos que representan a las páginas dentro del archivo. El nodo de las páginas a su vez está compuesto por información como el título, su *namespace*, su identificador y k nodos que representan cada una de sus revisiones con la información propia de estas.

› **Desarrollando un prototipo**

A partir de las métricas y de los medios de obtención de información que se definieron anteriormente se dio comienzo a un prototipo utilizando el lenguaje de programación java con librerías para el desarrollo de aplicaciones web. Se optó por una implementación con interfaz web que brinda flexibilidad al momento de elegir qué herramientas de visualización utilizar, por lo cual se alinea con la primera intención de este prototipo que es dar visibilidad gráfica a la información y las métricas generadas.

Este prototipo en java utiliza las tecnologías de *Spring MVC* e *Hibernate* que son descritos por Luna (2014). También se utilizó la librería *XStream* para facilitar el procesamiento de archivos XML (Frisch & Nakano, 2007: 2-13). El prototipo nos permite:

- Cargar *dumps* completos de *MediaWikis* a partir de archivos XML.
- Descargar y almacenar páginas individuales de la versión inglesa de *Wikipedia*.
- Obtener las categorías de las páginas descargadas de forma individual.
- Obtener estadísticas generales de la wiki al ser cargadas a partir de un *dump* completo.
- Obtener un listado de las páginas almacenadas.
- Por cada página obtener un listado de revisiones y poder ver las diferencias entre estas.
- Por cada página obtener de forma gráfica información sobre todas sus revisiones.
- Por cada página obtener de forma gráfica los cambios de estilos a lo largo del tiempo.
- Exportar información de los artículos en forma de *json* individuales.

A continuación, se detallan las visualizaciones realizadas por el prototipo para cada artículo.

Estadísticas de artículos

Aquí se presenta la información que obtenemos de la página de estadísticas de un artículo. El próximo gráfico representa el porcentaje de revisiones realizadas por cada usuario junto con un listado paginado de quienes se ocuparon de las ediciones (véase Figura 5). También podemos observar que al posicionarnos sobre alguna porción del diagrama circular se nos brinda información más detallada como, por ejemplo, la cantidad exacta de revisiones realizadas por dicho autor. Este gráfico es producto de aplicar la métrica *PercentOfRevisionByAuthor*, que a su vez utiliza las métricas *PercentOfAuthorRevisions* y *#RevisionsOfAuthor*, a la información de las revisiones de un artículo.

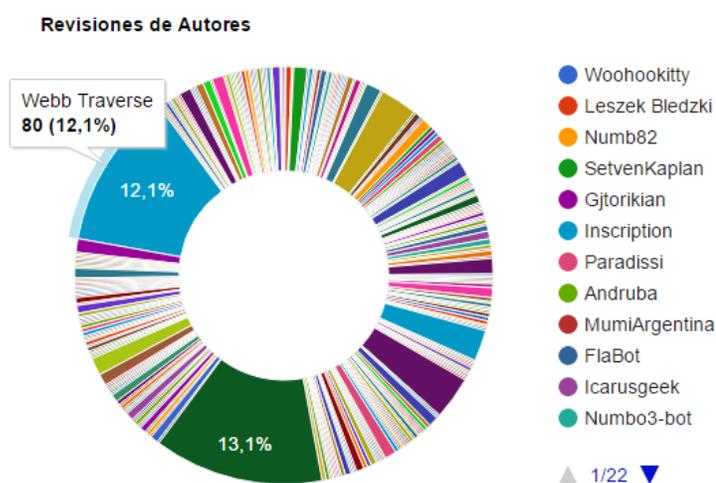


Figura 5. Revisiones de autores de un artículo.

Otro resultado de las estadísticas de revisiones se observa en una escala en la que se representa la cantidad de revisiones que se hicieron cada día (véase Figura 6). El mismo gráfico dispone de dos partes: la superior es un control que permite delimitar una porción de la información del total, que se representa en la parte inferior. En otra escala podemos ver esta funcionalidad en acción acotando el periodo de tiempo visualizado (véase Figura 7). También seleccionando parte de esta podemos obtener información específica de la cantidad de revisiones que se realizaron en una determinada fecha. La forma correcta de interpretar el gráfico es tomando el eje X como el periodo de tiempo desde que se creó el artículo hasta la actualidad, y el eje Y como el indicador del número de revisiones realizadas. Se demuestra, así, que cuanto más elevadas son las columnas de la escala, más revisiones se realizaron en un día en particular. Este gráfico fue realizado utilizando el resultado de aplicar la métrica *RevisionPerDay*, que a su vez emplea la métrica *RevisionsOfDay*, a la información de las revisiones de un artículo.

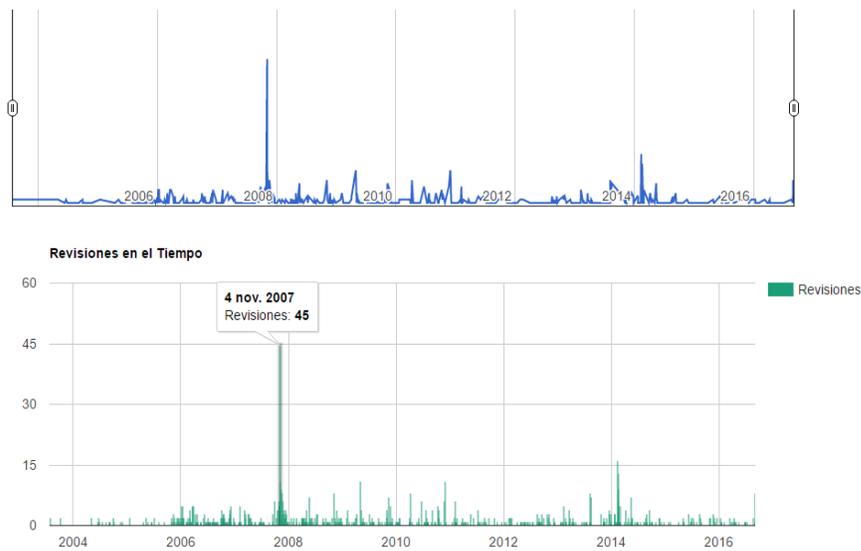


Figura 6. Revisiones en el tiempo de un artículo.

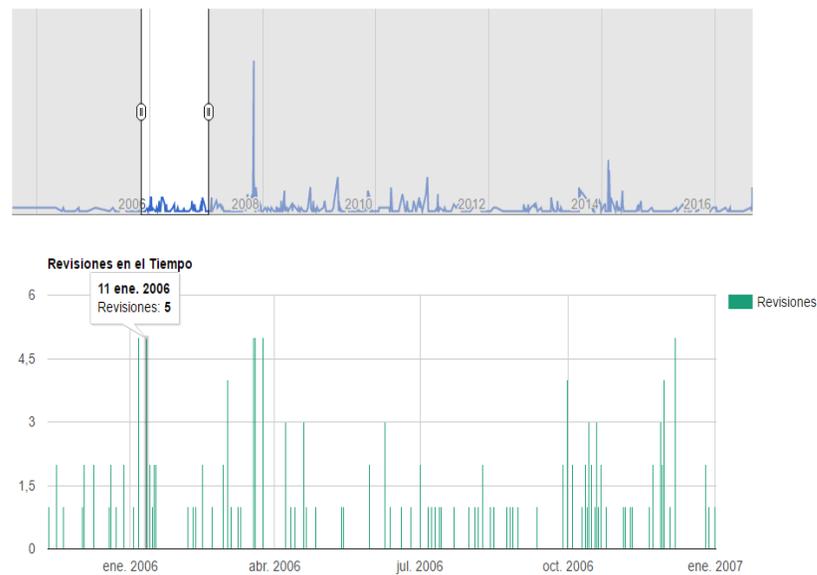


Figura 7. Revisiones en el tiempo de un artículo con zoom.

Para finalizar con las estadísticas de revisiones se presenta un gráfico en el que se muestra el tamaño en bytes de las revisiones en el tiempo (véase Figura 8). Se eligió esta representación por la relación de equivalencia entre un carácter y un byte. Este gráfico posee una visión de zoom al igual que en los anteriormente descritos y, al seleccionar parte de este, también se puede obtener información adicional como el tamaño exacto en bytes. Es

importante resaltar que en estas figuras pueden observarse líneas en blanco. No representan necesariamente que el contenido se haya reducido drásticamente a 0, sino una limitación del graficador el cual se alimenta de las revisiones para generar el esquema y, de no existir revisiones en esa fecha, simplemente la representa como si el contenido fuera de 0 bytes. En la escala 8 el eje X representa el periodo de tiempo desde que se creó el artículo hasta la actualidad, y el eje Y funciona como el indicador del número de bytes contenidos en el artículo. Se demuestra, entonces, que cuanto más elevadas son las columnas del gráfico, mayor es el tamaño del artículo para esa fecha. Este gráfico fue realizado utilizando el resultado de aplicar la métrica *BytesOfRevisionsPerDay* a la información de las revisiones de un artículo.

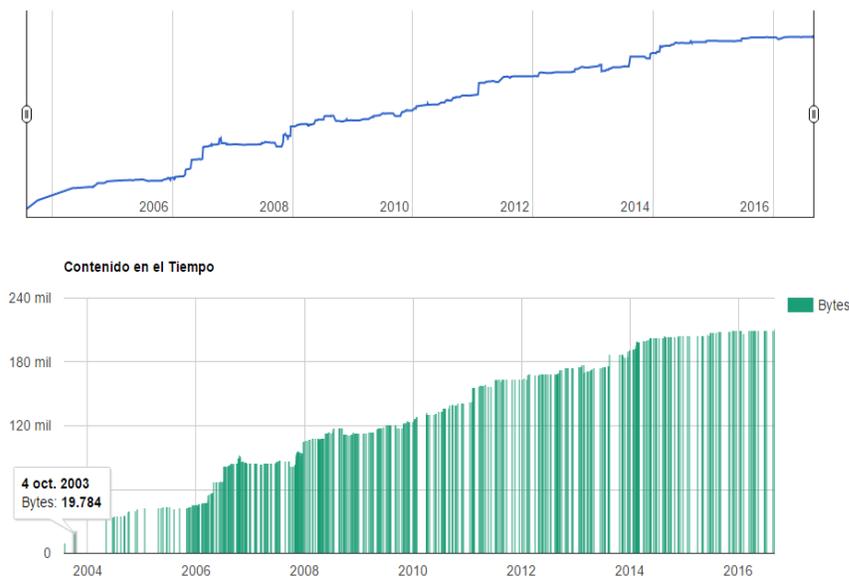


Figura 8. Contenido en el tiempo de un artículo.

Estadísticas de estilos de artículos

Continuando con las estadísticas se presentan las relacionadas con la información de los estilos. Un controlador en posición izquierda (véanse las Figuras 9 y 10) nos permite seleccionar cuáles tipos de estilos queremos visualizar en un gráfico. Podemos optar entonces por un modo individual (como en la Figura 9) o grupal (como en la Figura 10).



Figura 9. Estilos en el tiempo de un artículo. Solo Cursiva.

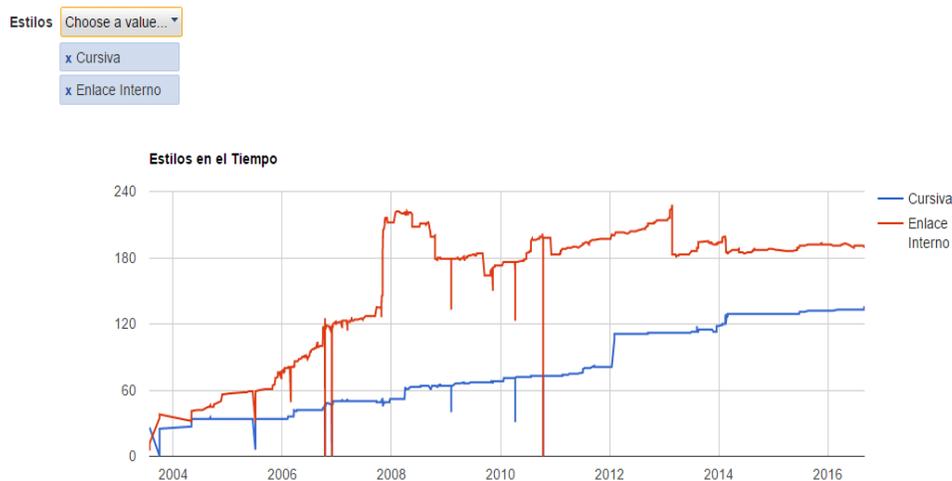


Figura 10. Estilos en el tiempo de un artículo. Estilos Cursiva y Enlace Interno.

A la derecha de la Figura 10 se presenta un listado que indica el color con el que se representa cada estilo en la escala; puede llegar a convertirse en un listado paginado según la cantidad de estilos presentados. Posicionándonos sobre ambos gráficos podemos obtener información adicional de la cantidad de apariciones de un estilo en particular. El eje X representa el periodo de tiempo desde que se creó el artículo hasta la actualidad, y el eje Y funciona como el indicador del número de aplicaciones del estilo en el artículo. Cuanto más elevados sean alguno/s de los trazos que representan a cada tipo de estilo seleccionado, mayor será la cantidad de sus aplicaciones para una fecha determinada. Las Figuras 9 y 10 fueron realizadas utilizando el resultado de aplicar la métrica `#OurrencesOfStyle` a la información de `markup` extraída del texto de cada revisión del artículo.

Por último, también podemos observar un gráfico (véase Figura 11) en el cual se visualiza la cantidad total de estilos por revisión.

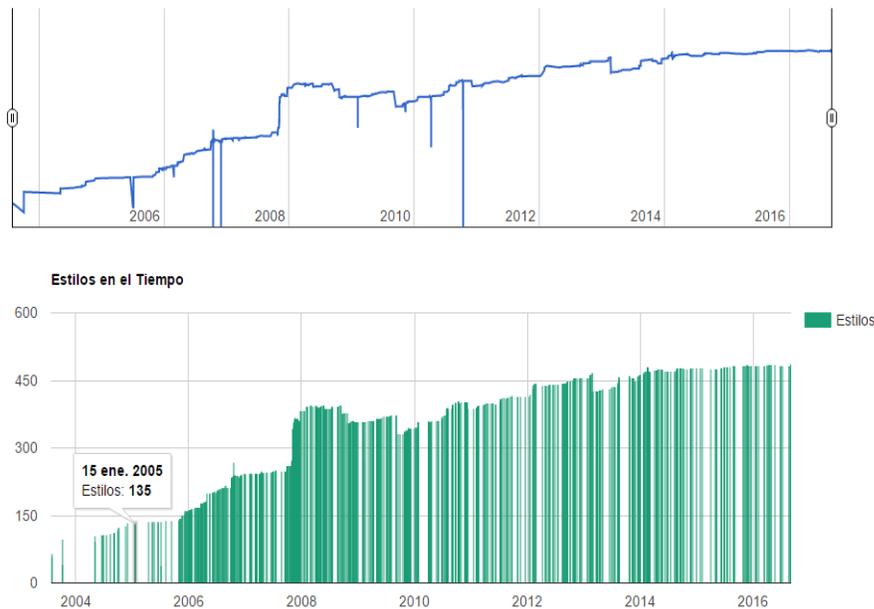


Figura 11. Estilos en el tiempo de un artículo. Estilos acumulados.

Al igual que en los gráficos anteriores, al posicionarse sobre este, se obtiene información sobre la cantidad exacta de estilos. Además, la escala provee del mismo tipo de funcionalidad de zoom utilizado en las anteriores. Su forma de lectura continúa la de los gráficos ya ilustrados, siendo el eje X el periodo de tiempo mientras el eje Y representa el valor a medir, este es, en esta ocasión, la sumatoria de todos los estilos de cada revisión. Cuando las columnas aumentan de tamaño, también aumenta el número de estilos aplicados en el artículo. Este diagrama fue realizado utilizando el resultado de aplicar la métrica *StylesOfRevisionsPerDay*, que a su vez emplea la métrica *StylesInText*, a la información de *markup* extraída del texto de cada revisión del artículo.

> **Evaluación y resultados**

Conjunto de datos

Como se mencionó anteriormente, en este trabajo se utilizarán wikis de la familia de *MediaWiki*. En particular, para realizar la evaluación se utilizaron las siguientes páginas extraídas de la versión en inglés de *Wikipedia*:

Título	Fecha de la primera revisión obtenida	Fecha de la última revisión obtenida	Cantidad de revisiones
Julio Cortázar	29 de julio de 2003	9 de agosto de 2016	1009
Pope	9 de noviembre de 2001	18 de septiembre de 2016	5374
Johnny Depp	14 de agosto de 2002	23 de septiembre de 2016	9638
Barack Obama	18 de marzo de 2004	22 de septiembre de 2016	25205

Tabla 2. Ejemplos de revisiones. Fuente: https://es.wikipedia.org/wiki/Wikipedia_en_ingles.

Análisis de resultados

Luego de analizar todos los resultados anteriores en busca de hitos o sucesos de interés que contengan relación con eventos históricos o circunstancias de importancia en las ediciones realizadas en *Wikipedia*, se encontraron varios hechos relevantes de los cuales algunos se detallarán a continuación.

Barack Obama

El 5 de noviembre de 2008 se detectó un incremento abrupto de la actividad de los editores en el artículo, el más alto en su historia (véase Figura 12).

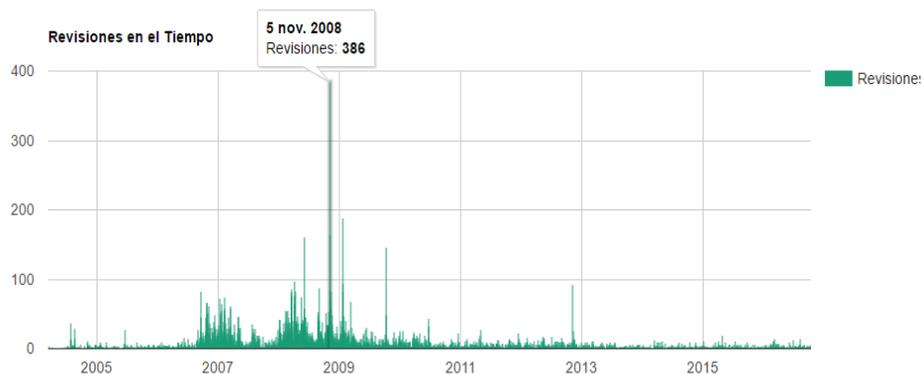


Figura 11. Revisiones en el tiempo de la página Barack Obama. 5 de noviembre de 2008.

También se detectó un incremento del tamaño del artículo, lo cual marca el comienzo de una nueva etapa de su crecimiento (véase Figura 13).

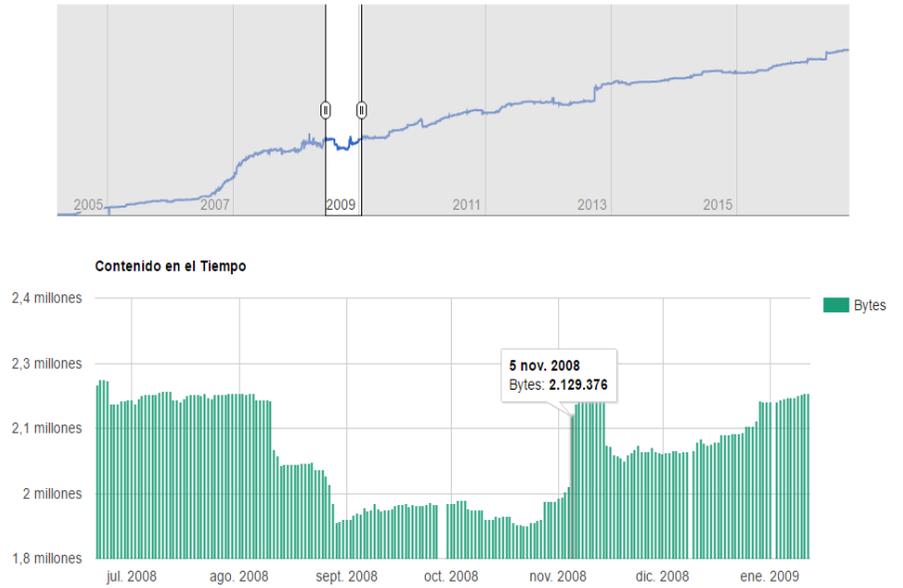


Figura 12. Contenido en el tiempo de la página Barack Obama. 5 de noviembre de 2008.

Por último, el siguiente gráfico contiene la suma de los estilos (véase Figura 14).

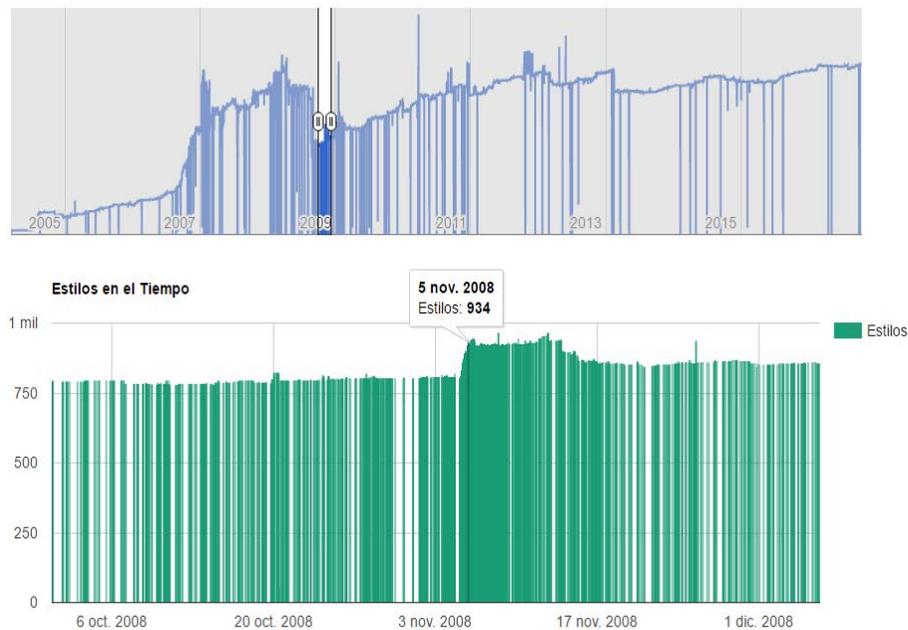


Figura 13. Estilos en el tiempo de la página Barack Obama. 5 de noviembre de 2008.

Según observamos en esta figura, se detectó un incremento de los estilos aplicados al artículo acompañando de igual forma el crecimiento del artículo. Esto se debió a que muchos editores comenzaron a la vez a realizar múltiples ediciones agregando y eliminando contenido

en busca de un balance frente a un crecimiento abrupto del artículo. Antes de la detección de este incremento de actividad ya se había registrado un periodo de crecimiento del artículo tanto en el nivel de la actividad como en el del tamaño y la cantidad de estilos aplicados. Esta etapa de crecimiento coincidió con el periodo electoral de los Estados Unidos, por lo cual se cree que el nivel de actividad detectado el 5 de noviembre de 2008 puede estar relacionado con la victoria de Obama en las elecciones presidenciales que tuvieron lugar el 4 de noviembre del mismo año.

Pope-Elecciones papales

En este caso, como podemos observar en el gráfico, las dos fechas en las que *Pope* tuvo sus mayores niveles de actividad coinciden con las fechas en las que fueron electos los papas Benedicto XVI, 19 de abril de 2005, y Francisco, 13 de marzo de 2013 (véase Figura 15).

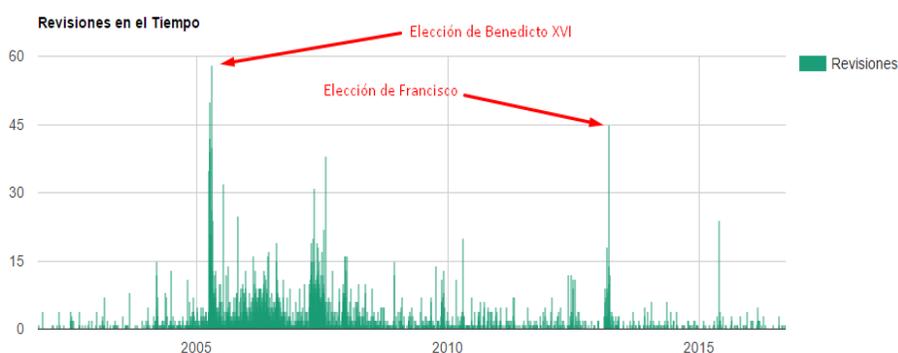


Figura 14. Revisiones en el tiempo de la página Papa (Pope). Elecciones papales.

Además, para cada una de esas fechas se evidencian cambios relevantes tanto en su contenido como en su estructura tal cual puede observarse en los dos gráficos siguientes, los cuales se corresponden con las fechas de la elección de los papas Benedicto XVI y Francisco, respectivamente (véanse las Figuras 16 y 17).



Figura 15. Estilos en el tiempo de la página Papa (Pope). 20 de abril de 2005.

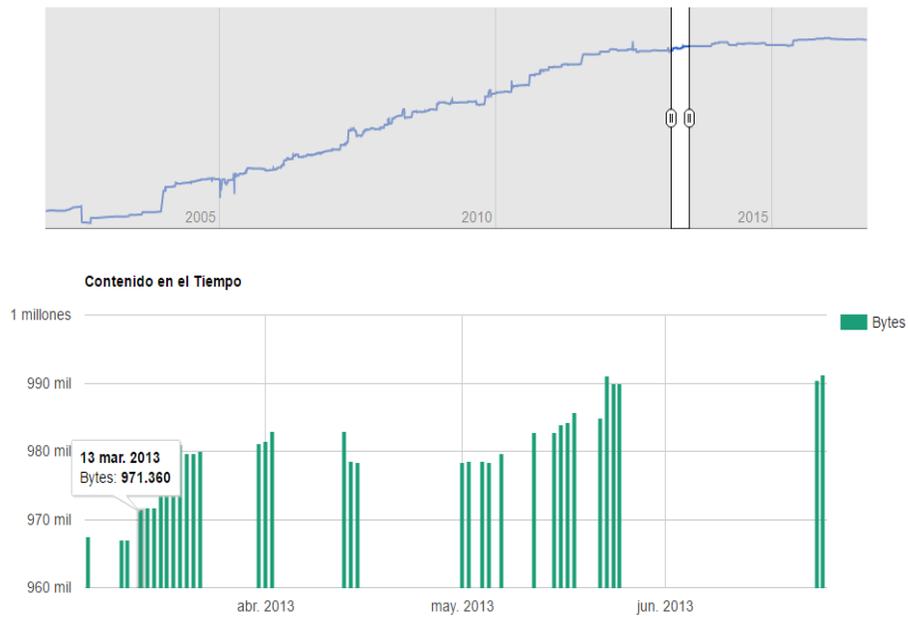


Figura 16. Contenido en el tiempo de la página Papa (Pope). 13 de marzo de 2013.

Por último, es importante aclarar que *Papa (Pope)* es una página referente al cargo, por lo cual su contenido varía según quién lo ejerza y es por ello que estos eventos desencadenaron una tendencia hacia la edición de este artículo.

Johnny Depp y Pope-Periodos de vandalismo

Referente a circunstancias de interés en las revisiones de los artículos de una wiki presentamos una relación entre nuestro trabajo y la detección automatizada de vandalismo.

El vandalismo en lo referente a ediciones en una wiki es toda revisión maliciosa o con intención de dañar el contenido del artículo. Esta se puede dar ya sea en forma de eliminaciones de contenido, ya sea agregando contenido contraproducente o a partir de la malversación del contenido que existe en un artículo.

Además, podemos observar en los próximos gráficos (véanse las Figuras 18 y 19), que representan la suma de estilos presente para cada una de las revisiones, periodos en los que se detectan múltiples caídas abruptas o picos aislados en la cantidad de estilos aplicados a los artículos.

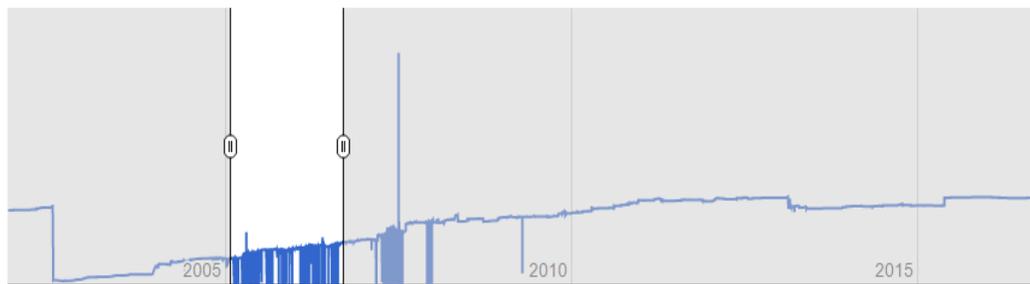


Figura 17. Estilos en el tiempo de la página Papa (Pope). Febrero de 2005 hasta agosto de 2006.

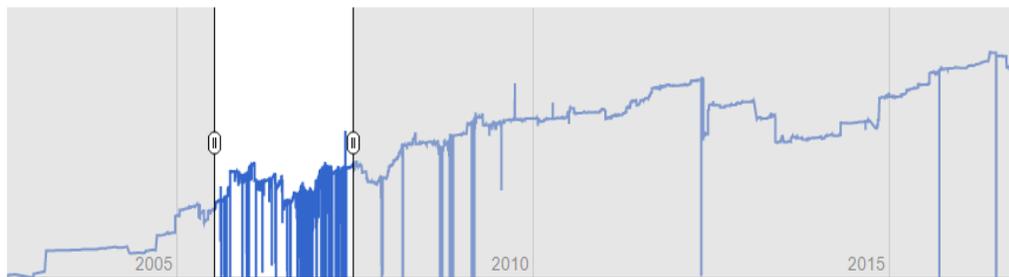


Figura 18. Estilos en el tiempo de la página Johnny Depp. Julio de 2005 a mayo de 2007.

Partiendo de la definición dada de vandalismo y de la comprobación de las revisiones en *Wikipedia*, en las cuales se observan casos en que el contenido es completamente eliminado o reemplazado por contenido sin sentido, podemos afirmar que esos fueron hechos de vandalismo que dieron lugar a periodos de vandalismo. Y estos suelen finalizar cuando se aplican a los artículos medidas de seguridad que previenen los actos de vandalismo.

También podemos observar que luego de los periodos de vandalismo aún ocurren casos en los que se registran caídas totales de estilos o estructura. A propósito de esto último,

consideramos que el análisis estructural o de estilos puede ser una variable de importancia para la detección de hechos de vandalismo en artículos.

› **Conclusión**

Este trabajo presenta un análisis referente a las fuentes de información y cómo esta se encuentra estructurada. Se analizan dos fuentes en profundidad: los *dumps* de *MediaWiki*, de los cuales se explica la forma de obtención y los diversos tipos de información contenida en estos, y también se estudia la obtención dinámica de los artículos de *MediaWiki* con sus respectivas revisiones. Por último, se plantean las ventajas y desventajas de cada fuente de información.

También se presenta un conjunto de métricas para evaluar la información obtenida y un enfoque basado en el lenguaje de marcado utilizado por las wikis para la organización, la estructuración y el formato de su contenido. A partir de estas métricas y de este enfoque particular se desarrolla un prototipo, el cual se aplica aquí a la *Wikipedia* en inglés para la obtención de información de las fuentes mencionadas.

El prototipo permite descargar y obtener la información sobre los artículos seleccionados, esto incluye toda la información de sus revisiones y su procesamiento para generar los valores utilizados en las métricas ya desarrolladas¹. Además de la obtención y del procesamiento de la información se producen gráficos referentes a la información recopilada. Estos permiten un análisis manual de los resultados.

A partir del prototipo también se presenta el análisis de un conjunto de datos acotados para posibilitar un análisis manual. Se presentan en primera instancia los resultados obtenidos directamente del prototipo y luego, un análisis de algunos de los casos de interés más relevantes detectados. Para finalizar se explica la importancia de estos resultados y su relación con los sucesos reales que los generan. Además, de los mismos se deriva un enfoque distinto para la detección de vandalismo en las revisiones a artículos de *Wikipedia*.

› **Trabajos futuros**

En este trabajo se utilizó un prototipo el cual obtenía parte de la información disponible dentro de las revisiones de las wikis. Una posible tarea para realizar es extender dicho prototipo para que abarque diferentes tipos de información como puede ser la extracción de contenido semántico de los artículos para ser analizado posteriormente. También el prototipo estaba orientado principalmente a artículos de wikis, y resulta de

¹ Véase el prototipo en el repositorio <https://github.com/jonx18/WebWiki>.

interés poder relacionar los cambios que suceden en estos textos con las páginas de discusión de los artículos en cuestión para poder obtener una mejor visión de la motivación de los cambios.

Como se planteó durante el trabajo, se espera que puedan generarse nuevas herramientas para la detección de vandalismo en artículos de *Wikipedia* y que estas posean entre sus variables de interés los cambios estructurales para mejorar la prevención de casos de vandalismo que aún los sistemas actuales no logran detectar.

A partir de los datos obtenidos por el prototipo se pretende generar mecanismos de análisis automático de la información para buscar patrones más complejos de los mencionados en este trabajo para comprender mejor la evolución de los artículos en las wikis. En caso de encontrarse estos tipos de patrones, también es importante categorizarlos para profundizar su comprensión.

Además, el objetivo mayor de todos los trabajos, incluyendo este, que abordan esta temática es generar una integración del estudio de los artículos, las revisiones, los editores y su actividad a fin de obtener un ecosistema en el que se tengan en cuenta todas las variables que influyen en la creación de conocimiento. Y de estas se podrá predecir la información relevante para el tipo de actividad desarrollada por un usuario en un momento determinado a fin de facilitarle y brindarle soporte a la tarea que se encuentre realizando.

› **Bibliografía**

Cunningham, W., Grigg, J. y Fletcher, J. (2014). Wiki History. Recuperado de <http://wiki.c2.com/?WikiHistory> el 22/12/2014.

Ejemplos de revisiones [Tabla]. Recuperado de https://es.wikipedia.org/wiki/Wikipedia_en_ingles el 20/02/2018.

Extracto del artículo "Clasificación TAS" de Wikipedia [Imagen digital]. Recuperado de https://es.wikipedia.org/wiki/Clasificaci%C3%B3n_TAS el 20/02/2018.

Frisch, A. y Nakano, K. (2007). *Streaming XML Transformation Using Term Rewriting. Programming Language Technologies for XML (PLAN-X)*, 2-13.

Historial de revisiones del artículo "Nupedia" [Imagen digital]. Recuperado de <https://en.wikipedia.org/> el 22/12/2014.

Luna, A. R. (2014). *Implementación del Patrón MVC en Aplicaciones Web con Java mediante la Integración de los Framework Hibernate, Spring y Primefaces*. Texcoco: Universidad Autónoma del Estado de México.

Menú para exportar páginas de Wikipedia [Imagen digital]. Recuperado de <https://en.wikipedia.org/> el 22/12/2014.

Página principal de Wikimedia [Imagen digital]. Recuperado de <https://dumps.wikimedia.org/> el 20/02/2018.