



¿Cómo y dónde editan las mujeres en las Wikipedias en español, catalán, euskera y gallego?

Julià Minguillón, Juan-José Boté

Wikimedia + Gendergap Research Meetup
Madrid - 01/04/2023



Universitat
Oberta
de Catalunya



UNIVERSITAT DE
BARCELONA

Facultat d'Informació
i Mitjans Audiovisuals



MINISTERIO
DE CIENCIA
E INNOVACIÓN



Tabla de contenidos

1. El proyecto: objetivos
2. Antecedentes
- 3. Metodología**
- 4. Ámbito del estudio**
5. Resultados
6. Conclusiones
7. Trabajo futuro

El proyecto

Objetivos

- Analizar la **participación** de l@s usuari@s registrad@s en las Wikipedias en español, catalán, euskera y gallego
 - Cuantificar la participación
 - Tipos de participación
- Analizar las diferencias en la participación en función del **género** registrado en el perfil de usuari@
 - Porcentaje de usuari@s registrad@s como mujeres
 - Diferencias en la participación, en cantidad y en tipología
- Analizar las diferencias entre Wikipedias

Antecedentes (I)

Objetivo: ¿cuántas mujeres editan Wikipedia y cómo y dónde lo hacen?

- Artículo en PLoS One
 - Wikipedia en castellano, hasta el 2017 incluido
 - Identificación del género
 - Mediante el perfil de usuari@ ⇒ mujer / hombre / no especificado
 - Mediante el análisis del contenido de la página de usuari@ (manual)
- Resultados
 - Hay pocas mujeres editoras, y editan menos y durante menos tiempo...
 - ... pero hay un subgrupo de mujeres tan activas como los hombres o más!

Antecedentes (II)

Objetivo: ¿por qué dejan o no de editar?

- Artículo en El Profesional de la Información
 - Investigación cualitativa, 12 mujeres entrevistadas de 142 contactadas
 - Diferentes niveles de participación
- Resultados
 - Algunas desconocían la posibilidad de especificar el género en su perfil de usuaria
 - El tópico sobre que las mujeres no saben editar es falso...
 - ...pero expresaron opiniones de inseguridad e inquietud para hacerlo muy bien
 - Abandono tras la finalización de Wikiproyectos si no se sabe qué continuar editando
 - Problemas con el uso de lenguaje sexista y el intento de usar uno inclusivo o neutro

Metodología

- 1) Obtener la participación de l@s usuari@s registrado@, usando los **volcados** (*dumps*) que se generan cada cierto tiempo
- 2) Determinar el género de cada usuari@ en función de lo que haya indicado en su **perfil de usuari@**
- 3) Para cada Wikipedia analizada, contar para cada usuari@ el número de ediciones que ha hecho en cada **espacio de nombres** (*namespace*)
- 4) Usar **métodos de agregación** para clasificar l@s usuari@s según su actividad en los diferentes espacios de nombres

Los volcados *¿qué?*

- Accesibles abiertamente, contienen una instantánea del estado de una Wikipedia en un idioma en un momento concreto

<https://dumps.wikimedia.org/eswiki/>

- Disponibles unos 8 volcados cada 2-3 semanas ⇒ decidir una fecha y descargar los volcados necesarios para el análisis (ENORMES!!!)
 - Actividad: el usuario *U* edita la página *P* en el momento *T* y hace los cambios *C*
 - Contenidos, categorías, ...

- Existen herramientas para manipular los volcados en Python

<https://pythonhosted.org/mediawiki-utilities/>

El perfil de usuari@ *¿quién?*

- Es posible registrarse en Wikipedia y editar de forma identificada (aunque anónima, bajo seudónimo)
- En el perfil de usuari@ es posible especificar el género: mujer / hombre / no especificado ⇒ Usuario: / Usuaría: (se puede consultar vía API)
- L@s usuari@s registrad@s pueden tener una página de usuari@ donde revelar información personal (entre otras cosas, su género)

Pregunta abierta: ¿Cuántas editoras tienen Usuaría: y/o página de usuaria?

Los espacios de nombres *¿dónde?*

- Todo lo que compone Wikipedia está organizado en [espacios de nombres](#)
 - Contenidos
 - Páginas (artículos)
 - Página de usuari@
 - Categorías
 - Wikipedia, Ayuda, Plantillas, ...
 - Wikiproyectos
 - ...
 - Cada tipo de contenido tiene una página de Discusión asociada

Problema: hay diferencias entre idiomas, p.e. anexos, Wikiproyectos, ...

El método de agregación *¿cómo?*

- Agrupación de espacios de nombres
 - Contenido
 - Discusión (todas)
 - Usuario
 - Wikipedia
 - Categoría
 - Wikiproyectos

- Parámetros del método de agregación
 - Usuari@s involucrad@s
 - Función de distancia
 - Algoritmo
 - Número de clústers

Ámbito del estudio

- Volcados de las [Wikipedias](#) castellana, catalana, gallega, y euskera (más la asturiana, la aragonesa y la extremeña)
- Datos desde 01/01/2018 hasta 31/12/2022 (**cinco años**)
- Usuari@s registrad@s con al menos **5 ediciones** en dicho período

	ES	CA	EU	GL	AST	AN	EXT
Volcado	9 GB	2.1 GB	528.8 MB	413.4 MB	241.5 MB	124.4 MB	8.0 MB
Ediciones	137 Mill.	30.3 Mill.	8.5 Mill.	6 Mill.	3.7 Mill.	1.8 Mill.	114 Mil
5 años	41.7 Mill.	11.5 Mill.	3 Mill.	1.6 Mill.	2.5 Mill.	280 Mil	10,407
Usuarios	124,996	12,303	5,955	2,008	681	371	101

Usuari@s registrad@s

- ¿Qué porcentaje de usuari@s especifica el género en cada Wikipedia?
- ¿Qué porcentaje son mujeres?

	ES	CA	EU	GL	AST	AN	EXT
Género	8.6%	10.8%	2.9%	10.6%	14.1%	11.3%	12.9%
Mujeres	18.6%	35.1%	21.7%	22.5%	10.4%	14.3%	23.1%

NOTA: las Wikipedias asturiana, aragonesa y extremeña tienen 10 o menos mujeres registradas con actividad suficiente en el periodo analizado

Pregunta abierta: ¿Por qué l@s usuari@s no identifican su género?

Resultados I (Wikipedia en español)

- Análisis de la participación (mediana) y retención

	Ediciones	Total días	En activo	1 día	1 semana	1 mes	6 meses
Mujeres	19	155	6	90.3%	81.4%	70.4%	51.1%
Hombres	26	819	9	95.4%	90.1%	85.2%	75.9%

- Presencia de usuari@s con decenas de miles de ediciones ⇒ ¿bots?
- Las mujeres dejan de editar a un ritmo mayor (el doble) que los hombres
- En las WP CA y GL es peor, en la EU es mejor (las mujeres editan más)
- Abandono en el caso de los hombres constante (\approx 20-25%), mujeres 2x, 3x

Pregunta abierta: ¿por qué no vuelven las mujeres a editar?

Resultados II (Wikipedia en español)

- Porcentaje de editor@s que participan (1 edición) por espacio de nombres

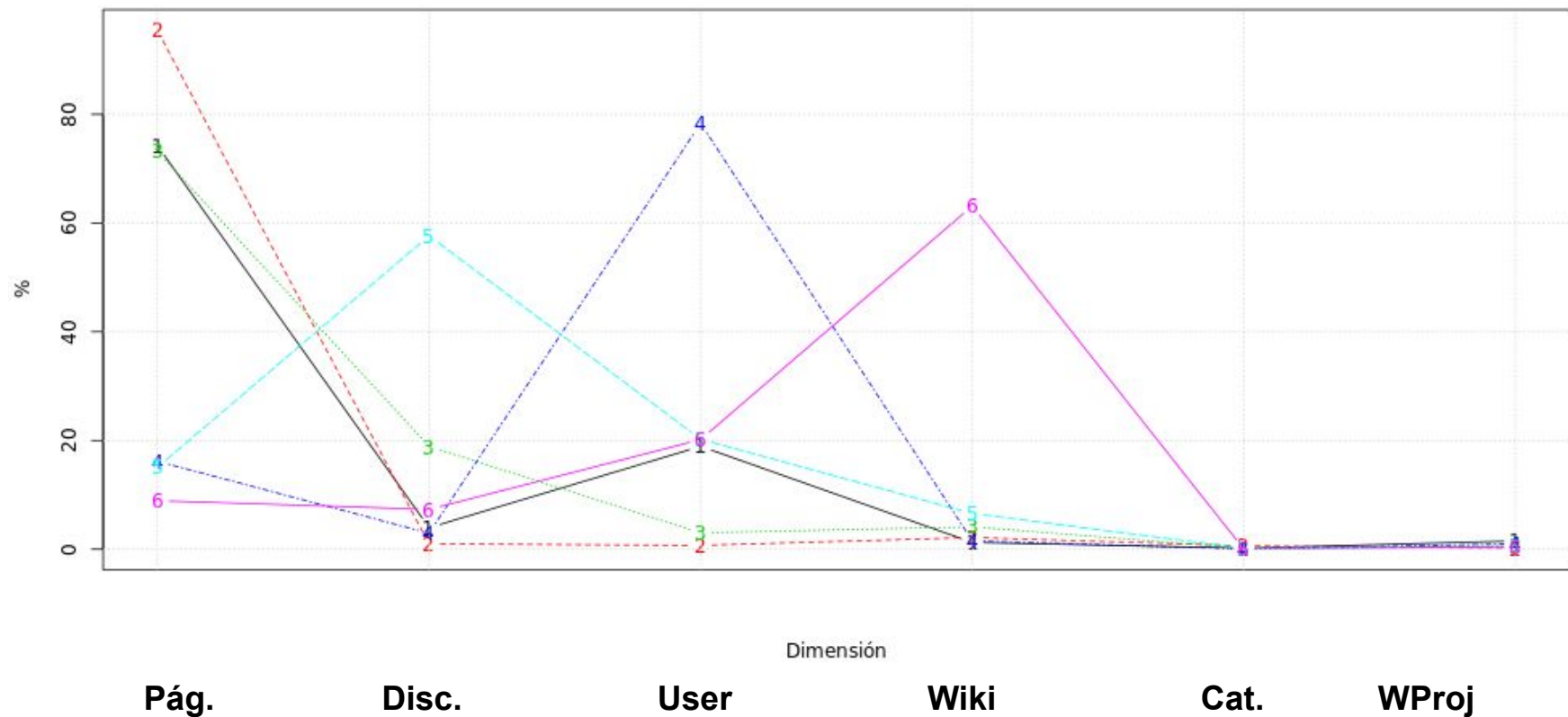
	Página	Discusión	Usuari@	WikiProy.	Categoría	Wikipedia
Mujeres	87.9%	42.3%	68.0%	15.6%	5.9%	32.5%
Hombres	95.2%	51.6%	49.1%	13.6%	12.9%	33.6%

- Pero, ¿y en porcentaje, cómo reparten su participación?

	Página	Discusión	Usuari@	WikiProy.	Categoría	Wikipedia
Mujeres	62.3%	7.0%	23.6%	1.2%	0.2%	5.7%
Hombres	77.8%	7.4%	10.4%	0.5%	0.4%	3.5%

Pregunta abierta: ¿Es mejor empezar editando solo la página de usuari@?

Resultados IIIa (Wikipedia en español)



Resultados IIIb (Wikipedia en español)

- Primeros resultados con 6 clústers (N=10,689, porcentaje de mujeres=18.6%)

	C1	C2	C3	C4	C5	C6
Usuari@s	2,439	4,927	1,721	870	474	258
% Mujeres / C	25.1%	12.2%	13.5%	39.9%	20.5%	37.2%
% Mujeres	30.8%	30.3%	11.7%	17.5%	4.9%	4.8%
% Hombres	21.0%	49.7%	17.1%	6.0%	4.3%	1.9%
Perfil	Cont. / User	Contenido	Cont. / mix	User / Cont.	Disc. / mix	Wiki / User

- Hipótesis: las mujeres pasan por fases, los hombres empiezan a editar ya
- Muy poca aportación de las dimensiones Categoría y Wikiproyecto

Conclusiones

- El porcentaje de mujeres editoras con perfil identificando su género es bajo, mostrando diferencias entre Wikipedias

Hipótesis: es posible que haya más mujeres, pero no lo revelan

- Respecto a estudios anteriores, el porcentaje parece haber aumentado, pero las editoras dejan de editar mucho antes

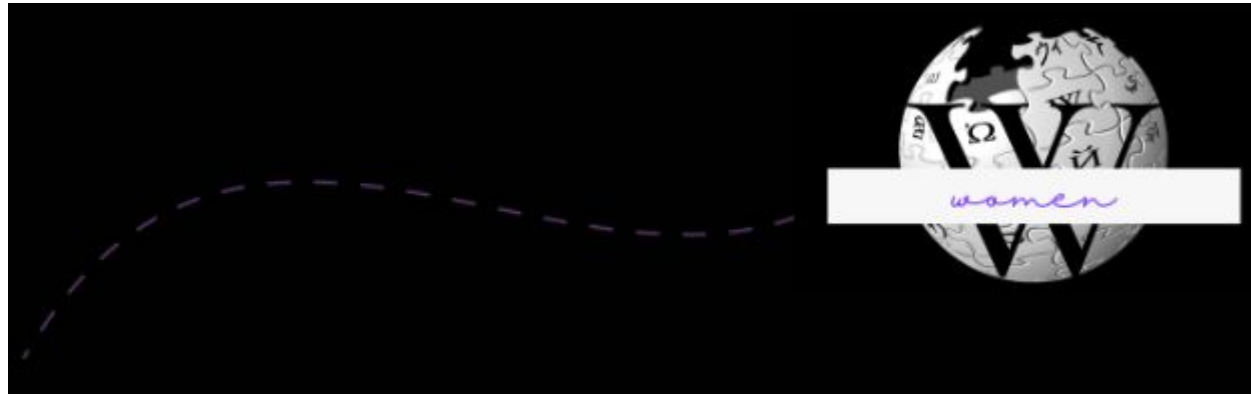
Hipótesis: siguen habiendo barreras después de captar nuevas editoras

- Se observan diferencias de comportamiento entre hombres y mujeres

Hipótesis: los hombres editan más por intereses, colaboran menos

Trabajo futuro

- Ajustar los parámetros del método de agregación
- Comparar incluyendo variables de género, antigüedad, número de ediciones, detalle de dónde editan, ...
- Comparar las diferentes Wikipedias entre sí
- Complementar el estudio con otros cualitativos en marcha



MUJERES y WIKIPEDIA: análisis de la brecha de género en la coproducción de conocimiento en las Wikipedias españolas

PID2020-116936RA-I00

<https://www.ub.edu/wikiwomen/>

Contacto: [jminguillona\[at\]uoc\[dot\]edu](mailto:jminguillona@uoc.edu)

