



# Escuela Superior de Ingeniería

Modelado, Simulación y Pruebas de Procesos y  
Tratamiento de Señales y de Datos

Trabajo de Investigación

## Estado del arte de la investigación sobre wikis

**Autor:** Emilio José Rodríguez Posada

**Tutor:** Juan Manuel Doderó Beardo

**Departamento:** Lenguajes y sistemas informáticos

Cádiz, Diciembre de 2012

# Índice

<b>1. Introducción</b>	<b>6</b>
<b>2. Motivación</b>	<b>7</b>
2.1. Páginas personales y blogs . . . . .	7
2.2. Gestores bibliográficos . . . . .	8
2.2.1. En servidores propios . . . . .	8
2.2.2. Servicios web y redes sociales . . . . .	8
2.3. Buscadores . . . . .	8
2.4. Recopilaciones en Wikipedia . . . . .	9
2.5. Revisiones de literatura . . . . .	9
<b>3. Objetivos</b>	<b>11</b>
<b>4. Definiciones, acrónimos y abreviaturas</b>	<b>12</b>
<b>5. WikiPapers</b>	<b>15</b>
5.1. Publicaciones . . . . .	15
5.2. Palabras clave . . . . .	16
5.3. Autores . . . . .	17
5.4. Datasets . . . . .	17
5.5. Herramientas . . . . .	17
5.6. Y más... . . . .	17
<b>6. Estado del arte</b>	<b>20</b>
6.1. Autoría y calidad . . . . .	20
6.2. Cobertura y sesgos . . . . .	22
6.3. Comunidad . . . . .	22
6.4. Educación . . . . .	23
6.5. Datasets . . . . .	24
6.6. GLAM . . . . .	24
6.7. Herramientas . . . . .	25
6.8. Motores wiki . . . . .	25
6.9. Predicción y tendencias . . . . .	26
6.10. Preservación . . . . .	26
6.11. Procesamiento del lenguaje natural . . . . .	27
6.12. Recomendación de tareas . . . . .	27
6.13. Semántica . . . . .	28
6.14. Vandalismo y spam . . . . .	29
6.15. Visualización . . . . .	30

6.16. Wikifarms . . . . .	31
<b>7. Cuestiones abiertas</b>	<b>40</b>
<b>8. Conclusiones y trabajo futuro</b>	<b>42</b>

## Índice de figuras

1.	Línea temporal de eventos sobre wikis. . . . .	7
2.	Portada de WikiPapers. . . . .	18
3.	Autores para los que se conoce su nacionalidad. . . . .	19

## Índice de cuadros

1.	Requisitos sugeridos en WikiSym . . . . .	16
2.	Datasets relacionados con wikis. . . . .	33
3.	Herramientas anti-vandalismo y anti-spam. . . . .	34
4.	Frameworks. . . . .	35
5.	Herramientas sobre lingüística. . . . .	36
6.	Herramientas de procesamiento de datos. . . . .	37
7.	Herramientas de visualización. . . . .	38
8.	Otras herramientas. . . . .	39

## 1. Introducción

El interés de los investigadores por los wikis, en especial Wikipedia, ha ido en aumento en los últimos años. La primera edición de WikiSym, un simposio sobre wikis, se celebró en 2005 y desde entonces han aparecido multitud de congresos, *workshops*, conferencias y competiciones en este área. El estudio de los wikis es un campo emergente y prolífico.

Ha habido varios intentos, aunque con escaso éxito, de recopilar toda la literatura sobre wikis. Unas veces el enfoque o la herramienta utilizada eran limitados, otras debido a las dimensiones de la tarea el proyecto era abandonado y al poco tiempo los metadatos bibliográficos se perdían. En este trabajo presentamos WikiPapers, un proyecto colaborativo para recopilar toda la literatura sobre wikis. Se hace uso de MediaWiki y su extensión semántica, ambos conocidos por los investigadores de este campo. Hasta noviembre de 2012 se han recopilado más de 1.700 publicaciones y sus metadatos, además de documentación sobre herramientas y *datasets* relacionados. Los metadatos son exportables en los formatos BibTeX, RDF, CSV y JSON. Los historiales completos del wiki están disponibles para descargar y facilitar su preservación. El proyecto está abierto a la participación de todo el mundo.

El resto del trabajo se divide de la siguiente manera. En la sección 2 motivamos este trabajo haciendo un repaso a los distintos enfoques utilizados hasta ahora para recopilar toda la literatura sobre wikis, incidiendo en sus ventajas e inconvenientes. En la sección 3 detallamos los objetivos. En la sección 4 definimos algunos términos que servirán para comprender mejor el contenido. En la sección 5 presentamos WikiPapers, cómo funciona y qué pasos se han dado. En la sección 6 hacemos un estado del arte empleando WikiPapers. En la sección 7 repasamos las cuestiones que a día de hoy siguen abiertas o que han tenido poca atención hasta ahora. Finalmente, en la sección 8, terminamos con unas conclusiones y trabajo futuro.

## 2. Motivación

El interés de los investigadores por los wikis, en especial Wikipedia, ha ido en aumento en los últimos años (Figura 1). La primera edición de WikiSym, un simposio sobre wikis, se celebró en 2005 y desde entonces han aparecido multitud de congresos (CLEF/PAN Lab), *workshops* (WikiAI, SemWiki y MathWikis), conferencias (Wikimania, WikiCon, SMWCon, Wiki Conference India, Wikipedia Academy y Wikipedia CPOV Conference) y competiciones (WikiViz). El estudio de los wikis es un campo emergente y prolífico. Ha habido varios intentos, aunque con escaso éxito, de recopilar toda la literatura sobre wikis (Ayers and Priedhorsky, 2011). Unas veces el enfoque o la herramienta utilizada eran limitados, otras debido a las dimensiones de la tarea el proyecto era abandonado y al poco tiempo los metadatos bibliográficos se perdían. Se han hecho recopilaciones en páginas personales y blogs, a través de revisiones de literatura, haciendo uso de gestores de bibliografía, en páginas de Wikipedia y también en servicios como Zotero o CiteULike. A continuación los describimos y evaluamos sus ventajas e inconvenientes y cómo WikiPapers resuelve las carencias de estos enfoques.

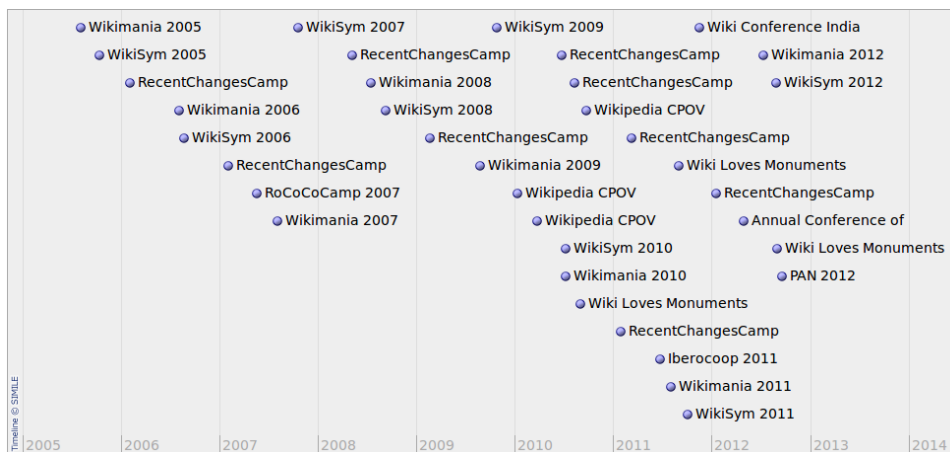


Figura 1: Línea temporal de eventos sobre wikis.

### 2.1. Páginas personales y blogs

Algunos autores han hecho recopilaciones de literatura en webs personales<sup>1</sup> y blogs. Un ejemplo bastante completo de este último tipo es SWEETpe-

<sup>1</sup><http://www.public.iastate.edu/~CYBERSTACKS/WikiBib.htm>

dia,<sup>2</sup> que contiene publicaciones sobre wikis y semántica. Uno de los inconvenientes de este sistema es que el esfuerzo suele recaer sobre una única persona y los metadatos no son fácilmente exportables. En WikiPapers el trabajo se hace colaborativamente y todos los metadatos son fácilmente exportables en diversos formatos.

## 2.2. Gestores bibliográficos

### 2.2.1. En servidores propios

Se han empleado gestores bibliográficos como WIKINDX<sup>3</sup> creando portales propios como Wikibibliographie ENCYCLEN<sup>4</sup> y otros ya desaparecidos, pero como contrapartida decidieron restringir la edición a un círculo de usuarios aprobados. Sin embargo, en WikiPapers pueden participar tanto usuarios registrados como sin registrar.

### 2.2.2. Servicios web y redes sociales

Existen servicios web y redes sociales con recopilaciones de literatura sobre wikis, en Zotero,<sup>5</sup> BibSonomy<sup>6 7</sup> y CiteULike.<sup>8 9 10</sup> Este enfoque sí hace uso de una comunidad de usuarios para procesar las publicaciones, pero de nuevo requieren registrarse, y la capacidad de estos servicios para aprovechar los metadatos generando tablas o gráficos es inexistente.

## 2.3. Buscadores

Google Scholar,<sup>11</sup> con buena cobertura, carece de representaciones visuales y no permite comentarios. Microsoft Academic Search<sup>12</sup> es gráficamente muy bueno, pero excluye material relacionado como datasets o herramientas. La capacidad de dar contexto a las publicaciones en WikiPapers de forma manual no puede ser superada por un buscador automático típico.

---

<sup>2</sup><http://www.mkbergman.com/sweetpedia/>

<sup>3</sup><http://sourceforge.net/projects/wikindx/>

<sup>4</sup>[http://wikindx.inrp.fr/biblio\\_encyclen/](http://wikindx.inrp.fr/biblio_encyclen/)

<sup>5</sup>[https://www.zotero.org/groups/wikipedia\\_research](https://www.zotero.org/groups/wikipedia_research)

<sup>6</sup><http://www.bibsonomy.org/tag/wikipedia>

<sup>7</sup><http://www.bibsonomy.org/tag/wiki>

<sup>8</sup><http://www.citeulike.org/tag/wikipedia>

<sup>9</sup><http://www.citeulike.org/tag/wiki>

<sup>10</sup><http://www.citeulike.org/group/382>

<sup>11</sup><http://scholar.google.es>

<sup>12</sup><http://academic.research.microsoft.com>



## 2.4. Recopilaciones en Wikipedia

También existen listados de publicaciones y recursos en algunas Wikipedias, como en la versión alemana<sup>13</sup> y la inglesa<sup>14</sup>. El principal inconveniente de este enfoque es que no es posible jugar con los datos dentro del mismo wiki, al estar todo escrito como texto plano, sin enriquecimiento semántico. En WikiPapers todos los metadatos son propiedades semánticas, tanto en las llamadas *infoboxes* (tablas) como en el cuerpo del artículo.

## 2.5. Revisiones de literatura

Finalmente, se han realizado varias revisiones de literatura con diferente grado de exhaustividad. La primera de ellas (Voss, 2005) se hizo en un momento en el que las publicaciones eran escasas, pero ya se podía ver una tendencia de publicación creciente y la presencia de muchas preguntas por responder. Un año más tarde (Ayers, 2006) vuelve a hacer un repaso a la literatura existente y enumera aquellas áreas que han recibido interés: historiales, páginas de discusión, contenido de los artículos, políticas del sitio, citas a artículos, encuestas a usuarios y listas de correo.

No sería hasta 3 años después cuando Okoli and Schabram (2009) presentan una propuesta de protocolo para hacer un mapeo sistemático de la literatura sobre Wikipedia, indicando la existencia de más de 1.000 publicaciones y ese mismo año Okoli (2009) analiza el estado del arte. Dos años más tarde, Nielsen (2011) hace la mayor revisión de literatura en un documento de más de 50 páginas, en progreso e inacabado, que incluye 300 referencias a publicaciones y reincide en la existencia de 1.000 publicaciones sobre el tema.

Martin (2011) hace primero un repaso técnico a la estructura de la base de datos (páginas, usuarios, texto), luego se centra en las áreas de calidad y acaba con unas consideraciones filosóficas.

Okoli *et al.* (2012) hacen una revisión de la literatura y han creado un wiki reutilizando las plantillas, formularios, semántica y estructura de WikiPapers. Están trabajando en el análisis de un subconjunto de la literatura, en este caso la relativa a Wikipedia exclusivamente. Es un proyecto temporal y pretenden incorporar sus resultados a WikiPapers.

La revisión más reciente Jullien (2012) hace un repaso por las motivaciones para contribuir, roles, estructura, la vida de un artículo, calidad,

---

<sup>13</sup><http://de.wikipedia.org/wiki/Wikipedia:Wikipedistik/Bibliographie>

<sup>14</sup>[http://en.wikipedia.org/wiki/Wikipedia:Academic\\_studies\\_of\\_Wikipedia](http://en.wikipedia.org/wiki/Wikipedia:Academic_studies_of_Wikipedia)

experiencia de usuario y accesibilidad, entre otros.

Uno de los inconvenientes de estas revisiones de literatura es que quedan rápidamente desactualizadas debido al ritmo con el que aparecen nuevas publicaciones. WikiPapers es actualizado continuamente por su comunidad de voluntarios.

### 3. Objetivos

Los objetivos de este proyecto de investigación son:

1. Idear un sistema o adaptar alguno existente que permita recopilar literatura sobre wikis, superando las carencias de los enfoques usados anteriormente.
2. Deberá permitir añadir, eliminar y modificar metadatos bibliográficos fácilmente, de manera manual e importando en lotes.
3. Usará características de la web semántica para poder sacar el máximo provecho posible a los metadatos.
4. Se pondrá atención en que los datos sean leibles tanto por humanos como por máquinas.
5. Tendrá licencia libre (por tanto los contenidos serán reutilizables por terceros) y facilitará la exportación de los metadatos.
6. Realizar un estado del arte haciendo uso del sistema creado en los puntos anteriores y con las publicaciones que se hayan podido recopilar hasta el momento. No será exhaustivo pero sí incluirá lo más relevante, ya que el propio proyecto es un estado del arte vivo que es continuamente actualizado.

## 4. Definiciones, acrónimos y abreviaturas

A continuación se incluyen algunas **definiciones, acrónimos y abreviaturas** que facilitan la comprensión de este trabajo.

**Administrador:** Persona que se dedica al mantenimiento de un sistema o red. En Wikipedia hace referencia a aquellos usuarios que pueden borrar/restaurar páginas, bloquear/desbloquear usuarios y proteger/desproteger páginas.

**Anónimo:** Usuario que no se ha registrado en el wiki y aparece identificado por su dirección IP.

**API:** Application Programming Interface. La API de MediaWiki proporciona acceso a los contenidos de las bases de datos.

**Artículo:** Página de un wiki situada en el espacio de nombres 0.

**Blanqueo:** Eliminación parcial o total del contenido de una página, lo que constituye una edición maliciosa o un acto de vandalismo.

**Bloqueo:** Suspensión temporal o indefinida a un usuario de su capacidad de modificar páginas. Los bloqueos sólo pueden realizarlos los administradores.

**Bot:** Programa que realiza tareas aburridas y tediosas de manera automática.

**Cambios recientes:** Página especial de los wikis en la que puede observarse las modificaciones realizadas.

**Cultura libre:** La componen todas las obras que tienen una licencia libre.

**Dataset:** Conjunto de datos con una estructura determinada que facilita su procesamiento.

**Diff:** Extracto que muestra las diferencias de contenido entre dos versiones distintas de una misma página.

**Discusión:** Anexo a cualquier página del software MediaWiki en la que se pueden discutir cambios en el contenido de dicha página.

**Edición:** Modificación de una página del wiki.

**Espacio de nombres:** División que realiza el software MediaWiki para diferenciar distintos tipos de páginas. Los principales son: artículos, categorías, plantillas y páginas de usuario.

**Etiqueta:** También conocida como netiquette, son una serie de normas de convivencia en comunidades en línea.

**Expresión regular:** Patrón que describe una o varias cadenas.

**Fork:** Bifurcación de un proyecto en dos distintos.

**GLAM:** acrónimo de Galerías, Bibliotecas, Archivos y Museos.

**Historial:** Conjunto formado por todas las versiones anteriores de una misma página, incluida la actual. Cada página mantiene su propio historial y es utilizado frecuentemente para restaurar el contenido debido a vandalismos, desacuerdos en la redacción, etc.

**Interwiki:** Vínculo que une a dos páginas sobre un mismo tema en distintos idiomas.

**Los cinco pilares:** Las cinco normas básicas de Wikipedia y en las que se sustentan el resto de políticas: (1) Wikipedia es una enciclopedia, (2) Wikipedia busca el punto de vista neutral, (3) Wikipedia es de contenido libre, (4) Wikipedia sigue unas normas de etiqueta, (5) Wikipedia no tiene normas firmes más allá de estos cinco pilares.

**Motor wiki:** software que facilita la redacción de documentos de manera colaborativa a través de una red.

**Namespace:** véase **Espacio de nombres**.

**NPOV:** véase **Punto de vista neutral**.

**Política:** Norma refrendada por la comunidad. Aunque Wikipedia no tiene normas firmes más allá de los cinco pilares, se recomienda seguir las políticas del proyecto, pues han sido elaboradas con un alto consenso y su no cumplimiento puede acarrear sanciones como bloqueos.

**Preservación digital:** tareas que permiten conservar datos digitales largo tiempo.

**Punto de vista neutral:** Es uno de los pilares de Wikipedia. Según el texto de la política oficial de Wikipedia en español: «El punto de vista neutral (PVN) establece que la enciclopedia debe contener hechos y que sus artículos deben ser escritos sin sesgos, presentando adecuadamente todos los puntos de vista existentes sobre tales hechos. [...] Esta política se malinterpreta con facilidad. No supone que sea posible escribir un artículo desde un único punto de vista objetivo no sesgado. Dice que debemos representar adecuadamente los diferentes puntos de vista y sin que el artículo afirme, implique o insinúe que alguno de ellos es el correcto. La neutralidad es mostrar todos los puntos de vista relevantes posibles tal y como son, para que cada lector adopte la opinión que prefiera».

**Regexp:** véase **Expresión regular**.

**Resumen de edición:** Texto que puede adjuntarse a cada modificación de una página con el propósito de explicar en qué consisten los cambios realizados. Es útil cuando se consulta el historial. Se considera una buena práctica el rellenarlo.

**Reversión:** Restaurar el contenido de una página a su estado anterior.

**Spam:** mensaje publicitario no deseado.

**Usabilidad:** define la facilidad de uso de una aplicación.

**Vandalismo:** Modificación no deseada y maliciosa en la que se elimina parte de la información, se introducen palabras soeces, etc.

**Web 2.0:** Término acuñado por Tim O'Reilly y que hace referencia a una nueva generación de la web, caracterizada por una mayor participación de los usuarios en los contenidos.

**Wiki:** Sitio web que permite a sus visitantes modificar el contenido del mismo. Ward Cunningham, desarrollador del primer software wiki, llamado WikiWikiWeb, lo definió como «the simplest online database that could possibly work».

**Wikifarm:** sitio que ofrece hosting para wikis.

## 5. WikiPapers

WikiPapers<sup>15</sup> fue lanzado en abril de 2011 (Figura 2). Haciendo uso de MediaWiki y su extensión semántica, recopila de manera colaborativa información acerca de toda la literatura sobre wikis, así como de herramientas y *datasets* relacionados. No hace falta estar registrado para participar, aunque es recomendable.

WikiPapers agrupa todas las ventajas de los sistemas mencionados anteriormente y soluciona sus inconvenientes. Permite hacer listados específicos de publicaciones, similares a SWEETpedia: por poner un ejemplo, existe uno de revisiones de literatura.<sup>16</sup> Funciona como un gestor bibliográfico, al almacenar los metadatos de las publicaciones y permitir hacer búsquedas, filtrarlos o exportarlos, individualmente o en conjunto. También facilita que grupos de usuarios se comuniquen a través de las páginas de discusión y compartan información sobre publicaciones de su interés, funcionando como una red social. Por otro lado, el espacio de discusión debajo de cada página posibilita a los lectores hacer valoraciones de los artículos. Y al ser un wiki, puede ser constantemente mejorado, evitando quedar desactualizado debido al rápido ritmo de publicación.

Desde un punto de vista más estadístico, es posible generar gráficas y tablas a partir de los metadatos disponibles en WikiPapers, aprovechando así la capacidad que ofrece la semántica. Gráficos de barras, circulares o líneas temporales están presentes y facilitan la visualización y comprensión de la información. También existe la posibilidad de incrustar diapositivas (SlideShare) y vídeos (YouTube, Vimeo).

Finalmente, el wiki y sus historiales están disponibles tanto para su descarga en XML y accesible a través de la API de MediaWiki. Esto impide que todo el trabajo se pierda, como ha sucedido en otros proyectos que quedaron inactivos y finalmente desaparecieron.

En WikiSym 2011 tuvo lugar un workshop (Ayers and Priedhorsky, 2011) donde se discutió acerca de qué requisitos debía cumplir un hipotético sistema que recopilara literatura sobre wikis (Cuadro 1). WikiPapers los cubre en su totalidad.

### 5.1. Publicaciones

En WikiPapers cada publicación dispone de una página en la que se detallan todos sus metadatos (título, autores, palabras clave, año, revista o

---

<sup>15</sup><http://wikipapers.referata.com>

<sup>16</sup>[http://wikipapers.referata.com/wiki/List\\_of\\_literature\\_reviews](http://wikipapers.referata.com/wiki/List_of_literature_reviews)

Característica	¿Lo ofrece WikiPapers?
Editable	Sí, es un MediaWiki
Listas/Grupos	Sí
Seguimiento/RSS	Sí
Ver páginas nuevas	Sí, ordenando por fecha
Categorías/Tags	Sí
Papers relacionados	Sí
Referencias/Citas	Sí
Conferencia/Journal	Sí, tesis y otros
Peer-reviewed o no	Sí, con un campo
Importar/Exportar	BibTeX/CSV/RDF/JSON
Usabilidad	Buena, Semantic Forms
API	Sí
Datasets	Sí
Enlaces	Sí
Trabajo futuro	Sí, preguntas abiertas
Free/Libre	Sí, excepto abstracts
Compatibilidad	RDF/Texto plano
Licencia de papers	Sí
Material relacionado	Sí
Stars/Comentarios	Sí
No wikitexto	Oculto, Semantic Forms

Cuadro 1: Requisitos sugeridos en WikiSym

congreso, DOI, idioma, licencia, enlaces al fichero y motores de búsqueda), el abstract, las referencias que incluye, las citas que recibe y un espacio de discusión. Los metadatos sirven para hacer búsquedas y filtrar. A noviembre de 2012 ya se dispone de más de 1.700 publicaciones,<sup>17</sup> incluyendo artículos de revistas y congresos, tesis y libros. Todos los metadatos se pueden exportar en los formatos BibTeX, RDF, CSV y JSON.

## 5.2. Palabras clave

Existe un listado de todas las palabras clave<sup>18</sup> presentes en los artículos y cada una de ellas cuenta con varios términos relacionados, lo que permite navegar entre ellas. Las más frecuentes son: Wikipedia, wiki, semantic

<sup>17</sup>[http://wikipapers.referata.com/wiki/List\\_of\\_publications](http://wikipapers.referata.com/wiki/List_of_publications)

<sup>18</sup>[http://wikipapers.referata.com/wiki/List\\_of\\_keywords](http://wikipapers.referata.com/wiki/List_of_keywords)



wiki, web 2.0, collaboration, evaluation, collaborative learning, knowledge management, MediaWiki, motivation, data mining y conflict.

### 5.3. Autores

Para cada autor hay una ficha que incluye su nombre, afiliación, país, índice de coautores, página web, estadísticas sobre número de publicaciones y citas, y por supuesto un listado de publicaciones, *datasets* y herramientas de su creación. Ya están listados unos 1.000 autores (Figura 3).<sup>19</sup>

### 5.4. Datasets

Un listado de *datasets*<sup>20</sup> permite observar la gran cantidad de datos sobre comunidades wiki disponibles para analizar. Existen *datasets* sobre vandalismo, texto wiki enriquecido con semántica, datos extraídos de *infoboxes*, *logs* anonimizados de visitas, mensajes de listas de correo y por supuesto los historiales de Wikipedia.

A este respecto, el proyecto WikiTeam<sup>21</sup> está compilando una gran cantidad de datos sobre comunidades wiki, que se cifra ya en 4.500 *dumps*.

### 5.5. Herramientas

Se está construyendo un listado de herramientas<sup>22</sup> que se han desarrollado a la hora de investigar sobre wikis y para proponer soluciones a ciertos problemas.

### 5.6. Y más...

También se está recopilando información sobre revistas, congresos, eventos, conceptos, ejemplos de análisis, preguntas abiertas, encuestas, motores wiki, wikifarms y más. WikiPapers, además de todo lo comentado anteriormente, sirve para que investigadores de wikis hagan comunidad y establezcan conexiones para futuras investigaciones.

---

<sup>19</sup>[http://wikipapers.referata.com/wiki/List\\_of\\_authors](http://wikipapers.referata.com/wiki/List_of_authors)

<sup>20</sup>[http://wikipapers.referata.com/wiki/List\\_of\\_datasets](http://wikipapers.referata.com/wiki/List_of_datasets)

<sup>21</sup><http://code.google.com/p/wikiteam/>

<sup>22</sup>[http://wikipapers.referata.com/wiki/List\\_of\\_tools](http://wikipapers.referata.com/wiki/List_of_tools)

Emijrp My talk Site settings Admin links My preferences My watchlist My contributions Log out

Page [Discussion](#) View [Edit](#) [History](#)

## Main Page

**WELCOME TO WIKIPAPERS**  
 WE ARE WORKING ON MORE THAN 2,925 PAGES WHICH HAD AROUND 220,000 VIEWS, AND YOU CAN HELP  
 ABOUT THIS WIKI | [NEW PAGES](#) | [LIST OF PUBLICATIONS](#) | [LIST OF AUTHORS](#) | [CATEGORIES](#) | [TUTORIAL ABOUT EDITING](#) | [HELP PAGES](#)

**WikiPapers** [\[edit\]](#)

WikiPapers is a compilation of resources (conference papers, journal articles, theses, books, datasets and tools) focused on the **research of wikis**. It aims to create the most comprehensive literature compilation for this research area. WikiPapers was started in April 2011 by Emilio J. Rodríguez-Posada, a pre-doctoral student at the University of Cádiz, Spain. [\(Read more...\)](#)

**How can I help?** [\[edit\]](#)

We are searching for volunteers. You can help filling missing fields in pages using the "Edit" tab or the (+) links. The following two categories include pages needing completion: [Category:Publication template parameter issues](#) and [Category:Author template parameter issues](#).

Also, you can create new pages and you are more than welcome to add your research publications regarding wikis. Finally, read WikiPapers, it is fun!

**How can I create a new page?** [\[edit\]](#)

For adding new content to WikiPapers, you can use the following forms:

- Form:Publication: books, conference papers, journal articles, etc.
- Form:Author: researchers.
- Form:Event: conferences, workshops, notable meetings, etc.
- Form:Keyword: keywords.
- Form:Dataset: datasets and corpuses.
- Form:Tool: tools and software.

There is a **TODO** list with many stuff to do. Be bold and join the effort! :)

**License** [\[edit\]](#)

About the **license** of this site:

- Abstracts are the property of their respective owners.
- Individual files have their own copyright terms.
- The rest of the wiki content is available under [CC-BY-SA 3.0](#) unless stated otherwise.

**Export** [\[edit\]](#)

You can **export** WikiPapers metadata in the following formats: RDF, CSV, JSON and BibTeX. Use the links available in the infoboxes. There is also an [API](#) which you can use to explore the wiki in an automatic way.

**Download** [\[edit\]](#)

You can **download a XML dump** with the full histories of this wiki. If this site vanishes in the future, nothing will be lost. Keep a copy on your computer. We use and recommend [WikiTeam](#) tools.

**Next events and calls for papers** [\[edit\]](#)

- Wikimania 2013, Hong Kong, August 7-11
- SMWCOn Fall 2012, Cologne, Germany, October 24-26
- PAN 2012, Rome, Italy, September 17-20

**Research areas** [\[edit\]](#)

Publications may be organized in the following areas: authorship • community, policies, social structure and organization • coverage and bias • datasets • history • image processing • literature reviews • motivations and incentives • quality • readership • recommender systems • semantic wikis • wikis as social networks • technical infrastructure • tools • usability • vandalism and spam • visualization

**Work in progress** [\[edit\]](#)

We are working on the following sections:

- List of affiliations (category)
- List of authors (category)
- List of countries (category)
- List of datasets (category)
- List of events (category)
- List of keywords (category)
- List of publications (category)
- List of tools (category)
- List of concepts
- List of examples
- List of open questions
- List of research areas
- List of journals (category)

**Publications distribution by year** [\[edit\]](#)

Year	Number of Publications
2001	0
2002	0
2003	0
2004	0
2005	10
2006	20
2007	130
2008	200
2009	230
2010	240
2011	240
2012	50

Publications, all types, distribution by year (See details)

Other academic wikis: [AcaWiki](#) • [Brede Wiki](#) • [WikiLit](#)

<< [WikiEvidens](#) • [WikiRing](#) • [WikiStats Extension](#) >>

Category: [Main categories](#)

This page was last modified on October 3, 2012, at 12:09.  
 This page has been accessed 15,523 times.  
 Abstracts are the property of their respective owners. Individual files have their own copyright terms. The rest of the wiki content is available under [Creative Commons CC-BY-SA 3.0 license](#) unless stated otherwise.

Privacy policy About WikiPapers Disclaimers

Figura 2: Portada de WikiPapers.

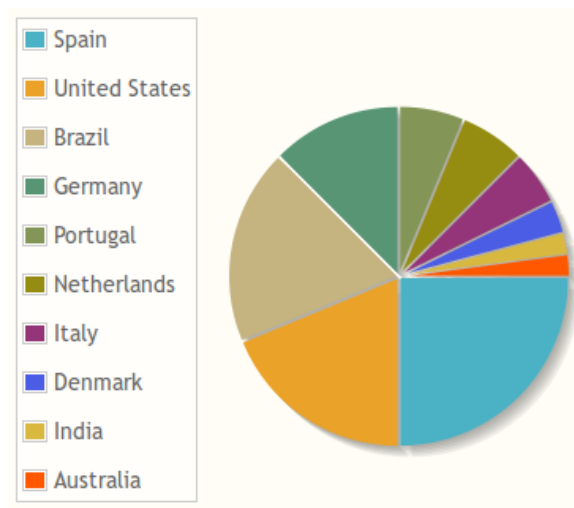


Figura 3: Autores para los que se conoce su nacionalidad.

## 6. Estado del arte

A continuación presentamos el **estado del arte** de la investigación sobre wikis haciendo uso de WikiPapers. No incluye todas las publicaciones posibles (está disponible en línea para consultas exhaustivas) pero sí los resultados más importantes hasta ahora, descartando experiencias duplicadas y otras que han quedado desfasadas. WikiPapers dispone actualmente de más de 1700 publicaciones sobre wikis que se han hallado de la siguiente manera:

- **Términos de búsqueda:** wiki, Wikipedia, Wiktionary, Wikibooks, Wikiversity, Wikiquote, Wikisource, Wikinews, Wikispecies, MediaWiki, Wikimedia, wikifarm.
- **Zonas de búsqueda:** título, abstract y palabras clave.
- **Servicios utilizados:** bases de datos (Scopus), buscadores (Google Scholar), redes sociales (Bibsonomy, CiteULike, CiteSeerX, Zotero) y repositorios de pre-prints (arXiv).
- **Fecha:** las búsquedas abarcan desde 2001 hasta 2012.

Ha requerido un esfuerzo de eliminación de algunos resultados no relacionados directamente con los wikis. Se escogieron los servicios enumerados por ser los más completos, aunque seguimos en búsqueda de otros y añadiendo las nuevas publicaciones que salen cada mes. En WikiPapers también se han creado páginas para los autores, *datasets* y herramientas relacionadas, a los que también les dedicamos una sección.

Agrupando las publicaciones por categorías temáticas hemos conseguido dividir el estado del arte en las siguientes áreas bien definidas.

### 6.1. Autoría y calidad

La naturaleza abierta de los wikis permite colaborar a todo aquel que lo desee (Leuf and Cunningham, 2001). Se pueden modificar sus contenidos simplemente registrándose en el sitio y muchas veces hasta de forma anónima, como en el caso de Wikipedia. Esto ha llevado a debates, a veces acalorados, sobre la calidad de los contenidos de los wikis.

Wikipedia no se presenta como un ente en posesión de la verdad, sino que expone los distintos puntos de vista<sup>23</sup> sobre cada tema indicando referencias a documentos, libros o webs externas que apoyan cada afirmación, dejando al

<sup>23</sup>[http://en.wikipedia.org/wiki/Wikipedia:Neutral\\_point\\_of\\_view](http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view)

lector que se forme una opinión contrastada, siempre existiendo la posibilidad de acceder a esas referencias para verificar la información. El proceso de añadir referencias es complejo, desigual en el tiempo y en él no participan todos los usuarios, sino un núcleo más comprometido con cada artículo (Chen and Roth, 2012).

Se han realizado comparaciones de calidad con formas más tradicionales de creación de obras. La más conocida y controvertida es la publicada en *Nature* (Giles, 2005) comparando entradas de Wikipedia con sus homólogas en la *Enciclopedia Britannica*. De 42 artículos analizados sobre ciencia, Wikipedia presentaba cuatro imprecisiones, mientras que *Britannica* tenía tres. Según *Nature*, *Britannica* poseía cierta ventaja sobre Wikipedia pero no tan grande como cabría esperar, al menos en cuanto a artículos científicos. Este estudio provocó que *Britannica* respondiese alegando que *Nature* había cometido errores en el análisis e interpretando sus propios resultados (Encyclopædia Britannica, 2006). *Nature* respondió a las acusaciones de *Britannica*, explicando la metodología y cómo había sido el proceso de revisión de los artículos, y mantuvo que el estudio era correcto (Nature, 2006).

Análisis como (Chesney, 2006) encontraron errores en un 13% de los artículos analizados y recogieron que los expertos otorgaban una mayor fiabilidad a los contenidos de Wikipedia mientras que los no expertos recelaban más. Otras comparaciones como (Rector, 2008) encontraron que los artículos de Wikipedia son fiables en un 80% frente al 95%-96% de sus homólogos en *Enciclopedia Britannica*, *The Dictionary of American History* y *American National Biography Online*.

Los mayoría de motores wiki mantienen los historiales completos desde que una página es creada. Esto permite conocer quién escribió qué y cuándo. Analizando los historiales es posible observar qué oraciones han perdurado más en el tiempo sin ser modificadas (Suzuki and Yoshikawa, 2012). Algunas herramientas, como WikiTrust (Adler *et al.*, 2008), usan estos y otros métodos para calcular indicadores de fiabilidad de los textos y lo visualizan con distintos colores (West and Lee, 2012). También se está trabajando en clasificadores de artículos por su calidad (Wu *et al.*, 2012), incluso se ha dedicado un congreso a este tema, llamado PAN 2012 Lab, del que han aparecido algunos de estos sistemas (Ferretti *et al.*, 2012) (Ferschke *et al.*, 2012). Iniciativas como Article Feedback Tool<sup>24</sup> ayudan a los lectores a evaluar la calidad de los artículos y a informar de errores.

La evaluación de la calidad de contenidos generados colaborativamente también está atrayendo interés en otras áreas, como en el caso de los mapas

---

<sup>24</sup>[http://en.wikipedia.org/wiki/Wikipedia:Article\\_Feedback\\_Tool](http://en.wikipedia.org/wiki/Wikipedia:Article_Feedback_Tool)

y callejeros de OpenStreetMap (Mashhadi *et al.*, 2012).

## 6.2. Cobertura y sesgos

El desarrollo de Wikipedia es muy desigual y existen problemas de **cobertura y sesgos** en los contenidos de cada idioma. Actualmente hay versiones de Wikipedia en más de 280 lenguas, siendo la inglesa la que cuenta con mayor número de artículos y usuarios. Wikipedia en inglés es la principal fuente desde la que otras ediciones traducen (Warncke-Wang *et al.*, 2012). Aunque temas básicos y fundamentales de arte, historia, geografía, ciencia, tecnología y ocio están presentes en una mayoría de idiomas, la cobertura de temas es diferente y está influenciada por la cultura de la que provengan sus usuarios (Massa and Scrinzi, 2012).

Se han desarrollado distintas experiencias con el ánimo de reducir esta brecha en los contenidos. Por ejemplo, se han invertido recursos en atraer a editores y lectores de los países que se encuentran en el denominado *Global South*<sup>25</sup>, que cuentan con una menor presencia en Wikipedia.

También se está presentando atención al número de mujeres que participan en los proyectos, intentando atraer un mayor número. Algunas herramientas como WikiTrip permiten observar el papel que juegan hombres y mujeres en el desarrollo de los artículos (Massa *et al.*, 2011). Algunos estudios indican que el número de aportaciones provenientes de mujeres oscilan entre el 5-20 % en función del idioma y proyecto. Hombres y mujeres hacen similar número de aportaciones, aunque sí se nota una mayor presencia masculina en el top de usuarios. (Antin *et al.*, 2011) Entre las hipótesis que se han dado para explicar la menor participación de mujeres es que en Wikipedia se producen muchos conflictos entre usuarios, que las mujeres rehuyen de estas situaciones y la falta de confianza en sí misma a la hora de editar el trabajo que otros han escrito (Collier and Bear, 2012).

## 6.3. Comunidad

A pesar de los sesgos comentados anteriormente, la **comunidad** de Wikipedia es una de las más ricas de Internet, por la presencia aunque en ocasiones sea minoritaria de personas de todos los países del mundo. En la versión inglesa por ejemplo, están registrados más de 17 millones de usuarios<sup>26</sup>, si bien es cierto que solo unos 130.000 han aportado algo en el último mes.

---

<sup>25</sup>[http://strategy.wikimedia.org/wiki/Wikimedia\\_Movement\\_Strategic\\_Plan\\_Summary/Increase\\_Participation](http://strategy.wikimedia.org/wiki/Wikimedia_Movement_Strategic_Plan_Summary/Increase_Participation)

<sup>26</sup><http://en.wikipedia.org/wiki/Special:Statistics>

Tal cantidad de usuarios, las interacciones que tienen, el contenido que generan y los conflictos que surgen, deben regularse de alguna forma. Aquí entran en juego los pilares de Wikipedia,<sup>27</sup> principalmente el de «Wikipedia está escrita bajo un punto de vista neutral», «es de contenido libre» y «compórtate civilizadamente», y una serie de normas y políticas<sup>28</sup> anexas para asuntos más específicos como el estilo de redacción o la resolución de conflictos.

El conocer cómo se organiza una comunidad y qué se espera de los usuarios nuevos es algo definitivo para la integración de los recién llegados. (Ford and Geiger, 2012) Estos primeros instantes y la convivencia posterior entre los usuarios definirá el éxito o fracaso en cuanto a retención y vida media de los mismos. (Zhang *et al.*, 2012)

La disminución de la retención de usuarios nuevos en los últimos años se ha achacado a que los recién llegados reciben un «bombardeo» de mensajes cuando actúan de modo incorrecto a las políticas de Wikipedia. Estos mensajes suelen estar ya escritos de antemano para ahorrar tiempo, son extensos, incluyen excesivos enlaces a páginas de normas y resultan muy fríos ya que no parecen escritos por un humano sino procedentes de algún sistema automático de avisos. Experimentos en los que se ha reducido el tamaño del mensaje, el número de enlaces y redactado de manera más informal han arrojado resultados positivos (Faulkner *et al.*, 2012).

## 6.4. Educación

Además de la utilidad en difusión de conocimiento que poseen intrínsecamente los wikis, también se les ha dado en **educación** un uso práctico para que los alumnos colaboren entre ellos, revisen los trabajos de sus compañeros y creen contenidos que servirán a terceros, frente al antiguo enfoque de hacer un trabajo que finalmente queda guardado en un cajón con un impacto mínimo.

Hasta ahora se han desarrollado experiencias docentes empleando wikis propios o wikis públicos como Wikipedia. En el primer caso se instala un wiki local y los alumnos trabajan con total libertad sobre él (Rodríguez-Posada *et al.*, 2011; Palomo-Duarte *et al.*, 2012a,c,b), en el segundo caso deben comprender las normas de Wikipedia y adaptarse a ellas. Este segundo enfoque es más delicado ya que si no se comprende el funcionamiento del proyecto los textos pueden acabar siendo borrados, algo que puede causar problemas al tratarse de trabajos para ser evaluados en una asignatura.

<sup>27</sup>[http://en.wikipedia.org/wiki/Wikipedia:Five\\_pillars](http://en.wikipedia.org/wiki/Wikipedia:Five_pillars)

<sup>28</sup>[http://en.wikipedia.org/wiki/Wikipedia:Policies\\_and\\_guidelines](http://en.wikipedia.org/wiki/Wikipedia:Policies_and_guidelines)

## 6.5. Datasets

Una de las ventajas de la investigación sobre wikis, sobre todo el caso de Wikipedia, es la disponibilidad de *datasets* con los textos completos, historiales, *logs* e información de actividad de los usuarios.

La mayoría de los wikis cuenta con algún tipo de licencia libre, generalmente Creative Commons o GFDL, lo que facilita el uso y distribución de los datos, aunque son pocos los que ponen a disposición del público copias completas de las bases de datos. Esto está cambiando a raíz de la aparición de WikiTeam,<sup>29</sup> un proyecto para desarrollar herramientas para preservar wikis que fundé en 2011 y que ya ha generado *datasets* para más de 5000 wikis con la ayuda de una comunidad muy comprometida.

Otros conjuntos de datos sobre wikis que están disponibles se encuentran en el Cuadro 2. Incluyen datos sobre coordenadas de artículos, semántica, páginas borradas, red de enlaces entre páginas, vandalismo, enlaces externos a repositorios, taxonomías y otros. La lista completa se encuentra en WikiPapers.<sup>30</sup>

## 6.6. GLAM

Las Galerías, Bibliotecas, Archivos y Museos o **GLAM** dedican mucho presupuesto a digitalizar material y a ponerlo a disposición del público en Internet. Análisis de los enlaces existentes hacia bibliotecas y archivos digitales de patrimonio cultural, realizados sobre Wikipedia en español y catalán, han mostrado que el número de referencias de este tipo es muy bajo aún para la mayoría de las colecciones, exceptuando algunas muy conocidas como la Biblioteca Virtual Miguel de Cervantes (Saorín-Pérez and Rodríguez-Posada, 2012).

La mayoría de estas instituciones han comprendido la importancia de que sitios tan visibles como Wikipedia, bien posicionados en buscadores como Google, incluyan enlaces a sus colecciones digitalizadas, para aumentar su tráfico y difusión (Lally and Dunford, 2007; Danielle Elder and Reilly, 2012). Tal es así que, por ejemplo, Europeana ha incluido un widget «citar en Wikipedia» para que sus contenidos sean fácilmente enlazables desde la enciclopedia libre. No se trata en exclusiva de dotar de visibilidad a estas colecciones digitales añadiendo estos enlaces a Wikipedia, sino de aumentar la calidad de los artículos de la enciclopedia, apoyando las afirmaciones que se hacen en los textos, los cuales muchos de ellos muestran deficiencias o son

---

<sup>29</sup><http://code.google.com/p/wikiteam/>

<sup>30</sup>[http://wikipapers.referata.com/wiki/List\\_of\\_datasets](http://wikipapers.referata.com/wiki/List_of_datasets)



referenciados con material poco académico (Årup Nielsen, 2007; Luyt and Tan, 2010).

Más cercano a la conservación y difusión del patrimonio monumental de los países se encuentra Wiki Loves Monuments. Esta iniciativa fue creada en 2010 con la intención de fotografiar monumentos para poder ilustrar los artículos de Wikipedia. La primera edición se realizó exclusivamente en los Países Bajos, en 2011 se extendió a 18 países europeos y en 2012 a otros países del mundo (Rodríguez-Posada *et al.*, 2012).

Para establecer vínculos entre la comunidad de editores de Wikipedia y estas instituciones culturales se ha creado la figura del «Wikipedian in Residence».<sup>31</sup> El wikipedista trabaja durante unos meses con la organización ayudándole a comprender las dinámicas de la enciclopedia libre, fomentando el desarrollo de los artículos que traten sobre las colecciones de la institución, organizando eventos o visitas donde se aprenda a participar en Wikipedia, entre otras posibles tareas como la coordinación en la liberación de material digitalizado bajo licencias libres compatibles con Wikipedia. Es una iniciativa que ha tenido bastante éxito y muchas instituciones culturales de todo el mundo se están acogiendo a ella con interés, por ejemplo el British Museum, Museu Picasso o Smithsonian Institution Archives.

## 6.7. Herramientas

Entorno a los wikis se generó un espectro de herramientas específicas. Unas (Cuadro 3) ayudan a detectar, reparar y eliminar vandalismos y ataques de spam, uno de los pocos inconvenientes que tienen los wikis al ser sistemas tan abiertos a la colaboración. Otras herramientas, más enfocadas a la investigación, facilitan el acceso a los datos de los wikis (Cuadro 4), a procesarlos (Cuadro 6) y visualizarlos (Cuadro 7). La lista completa se encuentra en WikiPapers.<sup>32</sup>

## 6.8. Motores wiki

Aunque MediaWiki sea el motor wiki más conocido y utilizado, hay muchos otros. La existencia de tantos motores dificulta la interoperabilidad entre ellos y existen intentos de unificar o al menos promover una sintaxis estándar, como el proyecto WikiCreole<sup>33</sup>.

---

<sup>31</sup>[http://outreach.wikimedia.org/wiki/Wikipedian\\_in\\_Residence](http://outreach.wikimedia.org/wiki/Wikipedian_in_Residence)

<sup>32</sup>[http://wikipapers.referata.com/wiki/List\\_of\\_tools](http://wikipapers.referata.com/wiki/List_of_tools)

<sup>33</sup><http://www.wikicreole.org>

Existe un portal que enumera a la mayoría de ellos, llamado WikiMatrix<sup>34</sup>. Contiene información acerca de más de 125 motores wiki y permite hacer comparaciones sobre la presencia o carencia de funcionalidades. Las características se organizan en las siguientes categorías: generales, requisitos del sistema, almacenamiento, seguridad y anti-spam, desarrollo y soporte, funcionalidades comunes y especiales, tipos de enlaces permitidos, sintaxis, usabilidad, estadísticas, formatos de salida, ficheros permitidos y funciones extra.

## 6.9. Predicción y tendencias

Wikipedia se ha constituido como una fuente de consulta de primer orden. Cuando ocurre algún suceso de gran envergadura (desastres, eventos deportivos, fallecimientos de famosos) inmediatamente se ve reflejado en el número de visitas y ediciones a los artículos relacionados.

Se han desarrollado sistemas que detectan estas tendencias y evalúan las subidas repentinas de actividad, mostrando los artículos que más interés han atraído en las últimas horas o días (Osborne *et al.*, 2012). También herramientas que evaluando las preferencias de los usuarios, les muestran aquellos temas candentes (Ciglan and Nørnvåg, 2010).

La magnitud de Wikipedia, la rapidez de actualización de sus contenidos y la gran cobertura que tiene para la práctica totalidad de temas de relevancia, la hacen una fuente de conocimiento estructurado de gran valor. Se están desarrollando algunos experimentos para utilizarla para predicciones, por ejemplo para prever recaudaciones en taquilla (Márton Mestyán and Kertész, 2012). Otros ejemplos incluyen el intento de intuir movimientos en los mercados financieros a través de Wikipedia.<sup>35</sup>

## 6.10. Preservación

Existe claramente un antes y un después en la generación y publicación de contenidos en Internet desde la invención del primer software wiki y de la creación de Wikipedia en 2001. Todo este contenido producido por millones de usuarios distribuidos por todo el planeta necesita ser preservado, no solo por su valor en conocimiento sino también por formar parte ya de la historia de Internet.

Proyectos wiki grandes y muy comprometidos con el conocimiento libre como Wikipedia publican copias de seguridad de todos sus artículos e histo-

---

<sup>34</sup><http://www.wikimatrix.org>

<sup>35</sup><http://rb.imm.dtu.dk/base/c/> y <http://rb.imm.dtu.dk/base/c/Lundbeck>

riales completos. Esto facilita la difusión y preservación para la posteridad. En cambio otros proyectos wiki más reducidos, pero muy numerosos, no tienen por costumbre publicar copias para descarga de sus contenidos.

Preservar los historiales completos y otros metadatos de los wikis no es una tarea trivial y plantea algunas dificultades. Por ello ciertas herramientas o servicios como Internet Archive o WebCitation no funcionan bien en estos casos. Curiosamente no ha habido mucho esfuerzo por cambiar la situación en este tema hasta hace relativamente poco tiempo.

(Popitsch *et al.*, 2010) propusieron un sistema para preservar los contenidos y metadatos de wikis que usaran distintos motores, de modo que también buscaban crear una sintaxis común. El proyecto quedó suspendido por falta de fondos.

Al año siguiente se creó WikiTeam<sup>36</sup>, un grupo de voluntarios que desarrolla herramientas de preservación para wikis. Hasta el momento funcionan únicamente sobre MediaWiki pero ya han logrado preservar más de 4.500 wikis de toda la red.

### 6.11. Procesamiento del lenguaje natural

Aunque Wikipedia dispone de un fondo audiovisual y más de 12 millones de imágenes, la parte fundamental la compone el texto de los artículos, 4 millones en el caso de Wikipedia en inglés. Esto compone un corpus inmenso sobre el que es posible aplicar técnicas de **procesamiento del lenguaje natural**.

La comparación del lenguaje utilizado entre artículos sobre desastres provocados por el hombre y aquellos causados por desastres naturales muestra que en los primeros son más frecuentes los términos relacionados con la ansiedad o ira, aún cuando se pretende que todos los artículos de la enciclopedia mantengan un punto de vista neutral y sin sesgos. (Ferron and Massa, 2012)

(Sateli and Witte, 2012)

### 6.12. Recomendación de tareas

La **recomendación de tareas** es un área todavía por explorar y explorar en los wikis. Dado que las comunidades wiki funcionan colaborando, los usuarios deben conocer qué artículos requieren su intervención para mejorar o ampliar los contenidos, subir imágenes, corregir errores, etc. Esta es una tarea que se hace de forma manual principalmente. El usuario recorre el wiki,

---

<sup>36</sup><http://code.google.com/p/wikiteam/>

leyendo sus contenidos, y descubre por azar páginas o secciones que necesitan ayuda. Hasta el momento existen muy pocas experiencias que hayan intentado resolver este asunto de forma automática.

Una de ellas es *Images for biographies*<sup>37</sup>, una aplicación web que muestra imágenes que pueden servir para las biografías de Wikipedia. Comparando las imágenes que se usan en otros idiomas de Wikipedia, recomienda esas imágenes para la misma biografía en otro idioma, si es que carece de fotografía. Este sistema ha ayudado a ilustrar miles de páginas y además es bastante sencillo de utilizar.

La gestión de errores en los artículos de Wikipedia siempre se había hecho utilizando las páginas de discusión adjuntas. En ellas la gente comenta los problemas que observan y se resuelven de manera conjunta. Estas páginas de informe de fallos pasan desapercibidas para los que son meros lectores de los artículos y no wikipedistas. Para implicar más a este sector de usuarios pasivos se han desarrollado algunos experimentos, por ejemplo Article Feedback Tool<sup>38</sup>, para facilitar el avisar de fallos. Es un trabajo en proceso pero está dando buenos resultados, ya que agiliza el informe y reparación de errores en los artículos.

Finalmente, se han desarrollado algunas metodologías y algoritmos para detectar artículos controversiales en base a cómo interactúan entre sí los usuarios que los redactan (Rad and Barbosa, 2012). Esto puede ayudar a encontrar páginas que requieran la participación de usuarios adicionales para resolver las disputas en los contenidos.

### 6.13. Semántica

La utilización de **semántica** en los wikis incrementa considerablemente la expresividad de los datos, permitiendo relacionar conceptos e inferir conocimiento. También permite a las máquinas leer lo que hasta ahora solo podían «comprender» los humanos.

El motor MediaWiki dispone de una extensión semántica llamada Semantic MediaWiki (Krötzsch *et al.*, 2006). Aunque inicialmente se desarrolló para ser añadida a Wikipedia, los requisitos de usabilidad y escalabilidad tan exigentes para el proyecto han hecho que a día de hoy aun no se haya incorporado a la enciclopedia libre. Para paliar esto se creó DBpedia (Auer *et al.*, 2007), que extrae datos estructurados de Wikipedia y los relaciona semánticamente. Otros wikis de Internet sí utilizan Semantic MediaWiki, por

---

<sup>37</sup><http://toolsserver.org/emijrp/imagesforbio/>

<sup>38</sup>[http://en.wikipedia.org/wiki/Wikipedia:Article\\_Feedback\\_Tool](http://en.wikipedia.org/wiki/Wikipedia:Article_Feedback_Tool)

ejemplo WikiPapers, e incluso existe una *wikifarm* llamada Referata<sup>39</sup> que ofrece wikis gratuitos con esta extensión semántica preinstalada.

#### 6.14. Vandalismo y spam

El carácter abierto de los wikis, en los que cualquiera puede modificar los contenidos, facilita la rápida mejora de los textos pero también que usuarios malintencionados o robots puedan causar destrozos, comúnmente llamados vandalismos, o añadir spam.

Se han empleado diferentes enfoques para contrarrestar estos efectos negativos. Se pueden clasificar en las siguientes categorías:

1. **Páginas de control de cambios:** se puede considerar el control más básico de todos, saber qué páginas han sido modificadas en los últimos instantes. Ejemplos: la página de «Cambios recientes».
2. **Herramientas semiautomáticas:** resaltan cambios que pueden ser dañinos, pero requieren intervención humana para decidir si revertir o no. Ejemplos: AbuseFilter, Huggle, STiki.
3. **Permisos de edición por grupo de usuarios:** en función de a qué grupo pertenece un usuario, podrá o tendrá vetada la modificación de ciertas páginas. Ejemplos: semiprotección y protección de páginas.
4. **Herramientas automáticas:** conocidas como bots anti-vandalismo. Existen bots especializados en mero vandalismo o en spam. Ejemplos: AVBOT, ClueBot, Salebot.
5. **Aprobación de cambios:** herramientas que acumulan los cambios y no los hace visibles hasta que alguien los aprueba. Ejemplos: FlaggedRevisions.
6. **Catpchas:** paran el spam antes de que se produzca, no quedando registrado en los historiales. Ejemplo: captcha de preguntas, reCAPTCHA.
7. **Soluciones específicas:** cuando hay muchas ediciones de un mismo usuario que revertir o muchas páginas que borrar debido a un ataque. Ejemplo: Rollbacker, la extensión Nuke.

Obviamente, las soluciones que requiere intervención humana son herramientas con nula o escasa «inteligencia» o algoritmia. Donde se ha desarrollado una mayor labor investigadora, aunque aun limitada, es en el desarrollo

---

<sup>39</sup><http://referata.com>

de herramientas automáticas que recorren los cambios recientes del wiki y revierten de manera autónoma cuando detectan un cambio dañino. Ejemplos de estos sistemas que llevan funcionando algunos años incluyen AVBOT (Rodríguez-Posada, 2009, 2010) en Wikipedia en español y ClueBot (Carter, 2008) en Wikipedia en inglés.

CLEF/PAN Lab es un congreso que ha centrado algunas de sus ediciones en la detección de vandalismos en wikis. Publicando un dataset con modificaciones marcadas como vandalismo y otras como benignas, invitaba a desarrollar algoritmos que aprendiendo de estos casos de entrenamiento, fueran capaces de detectar futuros casos de vandalismo.

El entrenamiento de algoritmos se basa como hemos comentado en datasets que contienen casos positivos y negativos. Lo que sucede es que es necesario que un humano haga esta categorización de antemano. Existe interés en desarrollar datasets utilizando los historiales de Wikipedia como fuente de casos positivos y negativos, a partir de las reversiones que hacen humanos (West *et al.*, 2010).

## 6.15. Visualización

La gran cantidad de datos disponible sobre wikis públicamente, además de que sus licencias permiten la libre reutilización, hacen necesaria y posible la aparición de numerosas herramientas de **visualización** que ayudan a entender cómo funcionan estas comunidades.

Desde los primeros años de Wikipedia se han creado herramientas y servicios para poder visualizar y comprender la evolución del proyecto. Desde tablas y gráficas sencillas sobre número de páginas, usuarios o ediciones por unidad de tiempo, hasta ranking de páginas según el nivel de actividad o mapas. Las estadísticas oficiales de Wikipedia se encuentran en stats.wikimedia.org y una recopilación con algunos ejemplos muy llamativos también está disponible.<sup>40</sup>

Entre las herramientas de visualización con mayor éxito en su momento fue HistoryFlow (Fernanda B. Viégas and Dave, 2004), que permite observar la evolución de un documento (en este caso una página cualquiera de Wikipedia) que haya sido redactado entre varios autores. Esto facilita ver qué partes de texto han perdurado en el tiempo, quién ha escrito qué oraciones, número de aportaciones, etc.

Para wikis con pocos usuarios y páginas StatMediaWiki<sup>41</sup> ofrece una gran cantidad de análisis. Mediante tablas y gráficas muestra la actividad por

---

<sup>40</sup><http://infodisiac.com/Wikimedia/Visualizations/>

<sup>41</sup><http://statmediawiki.forja.rediris.es>

usuario, página y grupos de páginas. Se ha utilizado para evaluar el trabajo de alumnos en wikis de asignaturas universitarias (Rodríguez-Posada *et al.*, 2011; Palomo-Duarte *et al.*, 2012a,c,b).

Para poder observar cuál es el ritmo de actividad de Wikipedia existen diversas herramientas, pero Wikistream<sup>42</sup> es una que cumple muy bien su cometido. Permite filtrar las modificaciones de usuarios registrados, anónimos y bots, también por idioma y espacio de nombres.

Existen muchas otras como *wmcharts*, *WikiTrip*, *WikiVis*, *Wikitweets*, etc. Cada año que pasa están disponibles nuevos conjuntos de datos y los que había con anterioridad crecen en tamaño. En 2011 tuvo lugar *WikiViz*, una competición que premiaba a los mejores trabajos que lograran visualizar el impacto de Wikipedia en la sociedad, lo que muestra el gran interés que hay en este sentido.

## 6.16. Wikifarms

El concepto wiki, como sitio web que permite a los usuarios consultar y modificar sus contenidos, lo creó Ward Cunningham en 1996. En 2001 se fundó Wikipedia y desde entonces esta forma de generar contenido colaborativamente se ha popularizado. El motor wiki más utilizado es MediaWiki, pero existen muchos más, entorno a los 100, la mayoría de ellos muy desconocidos.

Los usuarios de Wikipedia pronto se dieron cuenta de que, debido a las políticas de notabilidad y verificabilidad, había muchos contenidos que pudiéndose aprovechar del sistema wiki, no tenían cabida en la enciclopedia libre. Dado que MediaWiki es software libre, pronto aparecieron otros sitios web que haciendo uso de este motor, animaba a sus usuarios a generar contenido colaborativamente sobre los temas más variopintos.

Un paso más se dio cuando aparecieron las conocidas como **wikifarms** que no son más que *hostings* que facilitan al usuario la creación de sitios similares a Wikipedia, aislándote de todo el proceso de instalación en el servidor de las herramientas. En una wikifarm, con solo registrarte, dispones de un wiki listo para volcar material y mejorarlo con una comunidad.

Entre las wikifarms más conocidas se encuentra Wikia, pero existen otras como *EditThis*, *ShoutWiki*, *TropicalWikis* y *Wikikii*. También existe una especializada en wikis semánticos, *Referata*.

Uno de los inconvenientes de usar una wikifarm es que normalmente no se tiene acceso directo a la base de datos donde se almacenan todos los con-

---

<sup>42</sup><http://wikistream.inkdroid.org>

tenidos o al directorio donde se guardan las imágenes subidas. Esto supone, para el usuario externo, graves dificultades a la hora de hacer una copia de seguridad del wiki en cuanto tiene unos cientos de páginas e imágenes, ya que hace impracticable el guardar local y manualmente una copia de cada página.

Existen casos de wikifarms que han tenido problemas en el servidor, y dada la ausencia de copias de seguridad en el mismo y a la imposibilidad práctica de que los usuarios la hicieran para ellos mismos, han perdido todos los datos. En 2011 se fundó un proyecto llamado WikiTeam que ha desarrollado herramientas que recorren un wiki dado y descarga todos los contenidos, historiales e imágenes. Esta solución supone un gran avance en cuanto a la preservación de contenido wiki en Internet y reduce las inseguridades a la hora de utilizar una wikifarm para un proyecto propio.

En cuanto a la investigación de wikifarms, existe muy poco trabajo hecho hasta el momento. (Kittur and Kraut, 2010) hizo un análisis de 6000 wikis de Wikia, pero es poco frecuente ver estudios sobre otros wikis distintos a Wikipedia, cosa que debería ir cambiando con el tiempo si realmente se quiere tener una visión completa y no sesgada a la hora de analizar comunidades wiki de Internet.



Dataset	Tamaño	Idioma	Descripción
CoCoBi	5 MB	Alemán	Biografías de Wikipedia anotadas
Coordinates in Wikipedia articles	100 MB	Multilingüe	Artículos geocalizados de Wikipedia
DBpedia	1 GB	Multilingüe	Datos estructurados de Wikipedia
DeletionPedia	100 MB	Inglés	Artículos borrados de Wikipedia en inglés
Domas visits logs	40 GB/mes	Multilingüe	Número de visitas por página de Wikipedia
Google dataset linking strings and concepts	10 GB	Multilingüe	Relación entre enlaces y cadenas enlazadas
PAN Wikipedia vandalism corpus 2010	447 MB	Inglés	Ediciones marcadas como vandalismo
PAN Wikipedia vandalism corpus 2011	370 MB	Multilingüe	Ediciones marcadas como vandalismo
PlusPedia	33 MB	Alemán	Artículos borrados de Wikipedia en alemán
Repos-2012-dataset	5 MB	Multilingüe	Repositorios digitales enlazados desde Wikipedia
Social networks of Wikipedia dataset	100 MB	Multilingüe	Intercambio de mensajes entre usuarios
WikiBiography	11 MB	Alemán	Biografías de Wikipedia anotadas
WikiCorpus	1 GB	Multilingüe	Wikipedia enriquecida con lingüística
WikiIndex	100 MB	Inglés	Índice de wikis de Internet
WikiNet	100 MB	Multilingüe	Ontología extraída de Wikipedia
WikiPapers	10 MB	Inglés	Recopilación de literatura sobre wikis
WikiRelations	600 MB	Inglés	Relaciones entre categorías y páginas
WikiTaxonomy	2 MB	Inglés	Taxonomía a partir de Wikipedia
WikiTeam dumps	1 MB-10 GB	Multilingüe	Dumps de wikis de todo Internet
Wikia dumps	1 MB-10 GB	Multilingüe	Dumps de wikis de Wikia
Wikimedia dumps	1 MB-100 GB	Multilingüe	Dumps de wikis de Wikimedia
Wikipedia Historical Attributes Data	5 GB	Inglés	Atributos de plantillas <i>infobox</i> de Wikipedia
Wikipedia Vandalism Corpus (West)	25 MB	Inglés	Ediciones marcadas como vandalismo
Wikipedia page-to-page link database	400 MB	Inglés	Relación entre páginas de Wikipedia
WikipediaXML	1 MB-10 GB	Multilingüe	Dumps XML para information retrieval
Wlm-2011-dataset	5 MB	Multilingüe	Metadatos de imágenes de Wiki Loves Monuments 2011

Cuadro 2: Datasets relacionados con wikis.

Herramienta	S.O.	Idioma	Licencia	Descripción
AVBOT	Multiplataforma	Español	GPL	Bot anti-vandalismo para Wikipedia en español
ClueBot	GNU/Linux	Inglés	.	Bot anti-vandalismo para Wikipedia en inglés
CryptoDerk's Vandal Fighter	Multiplataforma	Inglés	Open source	Herramienta semi-automática anti-vandalismo
Huggle	Windows	Multilingüe	GPL	Herramienta semi-automática anti-vandalismo
Igloo	Multiplataforma	Inglés	Open source	Herramienta semi-automática anti-vandalismo
STiki	Multiplataforma	Inglés	GPL	Herramienta semi-automática anti-vandalismo
Twinkle	Multiplataforma	Inglés	.	Herramienta semi-automática anti-vandalismo
Vandal Fighter	Multiplataforma	Inglés	.	Herramienta semi-automática anti-vandalismo
VandalProof	Windows	Inglés	Freeware	Herramienta semi-automática anti-vandalismo
VandalSniper	Multiplataforma	Inglés	.	Herramienta semi-automática anti-vandalismo

Cuadro 3: Herramientas anti-vandalismo y anti-spam.

<b>Herramienta</b>	<b>S.O.</b>	<b>Idioma</b>	<b>Licencia</b>	<b>Descripción</b>
Java Wikipedia Library	Multiplataforma	Inglés	LGPL	Framework en Java.
Perlwikipedia	Multiplataforma	Inglés	GPL	Framework en Perl.
Python-wikitoos	Multiplataforma	Inglés	GPL	Framework en Python.
Pywikipediabot	Multiplataforma	Inglés	MIT license	Frame work en Python. El más usado.

Cuadro 4: Frameworks.

<b>Herramienta</b>	<b>S.O.</b>	<b>Idioma</b>	<b>Licencia</b>	<b>Descripción</b>
JWordNet-Similarity	Multiplataforma	Inglés	Gratis (uso no comercial)	Herramienta semántica para WordNet y Wikipedia
Manypedia.com	Multiplataforma	Inglés	Affero GPL	Comparación de Linguistic Points Of View de Wikipedias
Wikokit	Multiplataforma	Inglés	Varias	Framework para Wikcionarios
Zawilinski	Multiplataforma	Inglés	.	Extracción y análisis de gramática con Wikcionarios

Cuadro 5: Herramientas sobre lingüística.

Herramienta	S.O.	Idioma	Licencia	Descripción
DiffDB	Multiplataforma	Inglés	.	Trabaja con diffs de modificaciones en wikis
Ikiwiki	Multiplataforma	Inglés	.	Almacena un wiki en un repositorio git
Infobox2rdf	Multiplataforma	Inglés	GPL	RDFs a partir de parámetros de las <i>infoboxes</i>
MediaWiki API	Multiplataforma	Multilingüe	GPL	API para acceder a MediaWiki
Sioc MediaWiki	Multiplataforma	Inglés	.	RDFs a partir de artículos de Wikipedia
Wiki Edit History Analyzer	Multiplataforma	Inglés	.	Analiza cambios entre revisiones de una misma página
Wiki2XML parser	Multiplataforma	Inglés	.	Parser de dumps XML
WikiPrep	Multiplataforma	Inglés	GPL	Preprocesador de dumps XML
Wikipedia	Multiplataforma	Inglés	.	Map-reduce con Hadoop sobre dumps XML de wikis
Wikimedia Utilities	Multiplataforma	Inglés	MIT License	Lectura y extracción de los datos de dumps
Wikipedia Extractor	Multiplataforma	Inglés	GPL	Anota el contenido de los dumps
Wikipedia Miner	Multiplataforma	Inglés	.	Extrae estructura de enlaces, categorías, tipos...
Wikipedia-map-reduce	Multiplataforma	Inglés	Apache license	Análisis paralelo de las revisiones de un wiki

Cuadro 6: Herramientas de procesamiento de datos.

Herramienta	S.O.	Idioma	Licencia	Descripción
HistoryFlow	Multiplataforma	Inglés	.	Visualización de la evolución de un texto
StatMediaWiki	Multiplataforma	Inglés	GPL	Estadística y visualización de wikis MediaWiki
Wiki Explorer	Multiplataforma	Inglés	.	Biblioteca para exploración y visualización de CMS (MediaWiki)
Wiki Trip	Multiplataforma	Inglés	GPL	Visualización de la evolución de páginas de Wikipedia
WikiBlame	Multiplataforma	Multilingüe	.	Encuentra cuándo y quién añadió cierto contenido a un wiki
WikiChanges	Multiplataforma	Inglés	.	Líneas temporales interactivas de actividad en wikis
WikiEvidens	Multiplataforma	Inglés	GPL	Estadística y visualización de wikis MediaWiki
WikiPride	Multiplataforma	Inglés	GPL	Preprocesado y visualización de dumps MediaWiki
WikiScanner	Multiplataforma	Inglés	.	Muestra qué empresas y organismos editan Wikipedia
WikiTracer	Multiplataforma	Inglés	.	Visualización para varios motores wiki
WikiTrust	Multiplataforma	Inglés	New BSD/GPL	Cálculo de reputación de textos
WikiVis (FH-KL)	Multiplataforma	Inglés	GPL	Visualización de actividad de usuarios, páginas, categorías
WikiVis (UM)	Multiplataforma	Inglés	ECL	Visualización de actividad de usuarios, páginas, categorías
WikiWarMonitor	Multiplataforma	Inglés	.	Medición de guerras de ediciones
WikiXRay	Multiplataforma	Inglés	GPL	Análisis cuantitativo
Wikistream	Multiplataforma	Inglés	CC0	Streaming de ediciones en Wikipedia
Wmcharts	Multiplataforma	Inglés	GPL	Estadísticas usuarios, páginas, actividad global

Cuadro 7: Herramientas de visualización.

<b>Herramienta</b>	<b>S.O.</b>	<b>Idioma</b>	<b>Licencia</b>	<b>Descripción</b>
WikiSim	Multiplataforma	Inglés	University of Edinburgh GNU license	Simulador de colecciones de conocimiento.
WikiTeam tools	Multiplataforma	Inglés	GPL	Preservación y backups de wikis.

Cuadro 8: Otras herramientas.

## 7. Cuestiones abiertas

Es cierto que algunas áreas en el campo de los wikis han sido muy prolíficas en cuanto a publicaciones, pero aun quedan muchas **cuestiones abiertas**, incluso algunos temas están mínimamente explorados.

Uno muy poco desarrollado y que podría tener un efecto muy positivo es el de los **sistemas de recomendación**. Fabricar sistemas que indiquen a los usuarios dónde es necesaria su intervención en función de sus preferencias sería muy beneficioso. Por ejemplo, en el contexto de Wikipedia, automatizar la petición de imágenes para artículos sobre municipios o monumentos según la localización del usuario o sugerir artículos a mejorar que estén relacionadas con los temas que el usuario lee o edita con frecuencia. O también detectar temas que falten, utilizando otras Wikipedias como fuente, ya que el desarrollo de las enciclopedias en distintos idiomas es desigual, excepto en temas básicos, ya que lo que exista en una no tiene porqué existir en otras más lejanas (por diferencias culturales). Esto ayudaría a optimizar la participación y a eliminar la falsa sensación de que «en Wikipedia está casi todo hecho ya».

Otra cuestión a explorar es el **análisis de imágenes**. Wikimedia Commons es una mediateca libre con más de 12 millones de imágenes y su mantenimiento es muy costoso de hacer manualmente. Herramientas que pudieran ayudar a organizar las imágenes reconociendo objetos o características destacables liberarían de bastante trabajo a los miembros del proyecto.

Un asunto importante es que la inmensa mayoría de publicaciones hablan exclusivamente de Wikipedia y más concretamente de la versión inglesa, **ignorando sus proyectos hermanos u otros idiomas**. El segundo proyecto wiki más estudiado es Wiktionary, pero otros como Wikiversidad, Wikilibros o Wikinoticias han atraído escasa o nula atención. Y no solo eso, sino también **se obvian las miles de comunidades wiki** existentes en la red. No se reproducen los estudios o experiencias en estos otros lugares, donde quizás puedan salir resultados distintos o corroborarlos. Es cierto que hasta hace poco no era sencillo hacerse con una copia completa de las páginas, historiales y metadatos de la mayoría de los wikis de Internet para analizarlos, a menos que el administrador colaborara proporcionándola. Con WikiTeam esto está cambiando, ya que sus herramientas extraen todos los datos de manera sistemática sin necesidad de intervención de los administradores de cada wiki, lo que abre nuevas posibilidades.

En cuanto al estudio de **algoritmos para luchar contra el vandalismo**, si bien se han hecho avances y ha habido congresos sobre el tema, distan de tener una precisión adecuada para funcionar de manera autónoma,



a menos que se restrinja mucho su actuación solo a los casos de vandalismo o spam más flagrantes. Aun queda mucho por mejorar en este sentido.

Para finalizar, recientemente se ha creado **Wikidata**, un nuevo proyecto hermano de Wikipedia que recopilará una ingente cantidad de datos de forma estructurada que luego serán utilizados en Wikipedia y otros proyectos. Es muy posible que genere interés entre los investigadores en cuanto adquiera algo de tamaño, como sucedió en el pasado con la enciclopedia libre.

## 8. Conclusiones y trabajo futuro

En este proyecto de investigación hemos analizado los distintos enfoques utilizados hasta ahora para recopilar literatura sobre wikis. Tras evaluar sus ventajas e inconvenientes, procedimos a desarrollar un sistema, que solucionara las deficiencias conservando lo que se ha comprobado que funciona bien en otros casos y añadiendo las sugerencias que investigadores habían comentado en distintos foros. Entre las características fundamentales de este nuevo enfoque se encuentran: permite participar en la recopilación de literatura de manera colaborativa, aprovecha los metadatos para convertirlos en propiedades de la web semántica (relaciona conceptos, permite consultas, infiere conocimiento), facilita la continua actualización y puesta al día de los contenidos evitando quedar desactualizado como los Systematic Literature Review típicos y hace trivial la exportación y reutilización de los datos, entre otros.

El proyecto resultante ha sido WikiPapers, un wiki para recopilar literatura sobre wikis e información sobre *datasets* y herramientas relacionadas. Hasta el momento se han añadido metadatos para más de 1700 publicaciones, 50 datasets y casi 100 herramientas. La literatura sobre wikis se ha clasificado en las siguientes categorías: autoría y calidad, cobertura y segos, comunidad, educación, datasets, GLAM, herramientas, motores wiki, predicción y tendencias, preservación, procesamiento del lenguaje natural, recomendación de tareas, semántica, vandalismo y spam, visualización y wikifarms. El estado del arte planteado en este trabajo sigue esta organización.

Aunque WikiPapers ya ha conseguido recopilar una gran parte de la literatura sobre wikis, aun es posible mejorar la cobertura y seguir añadiendo las últimas publicaciones que vayan saliendo. Para esto es necesario conformar una comunidad comprometida con el proyecto. Varios investigadores de wikis ya han mostrado interés y participado, incluyendo sus publicaciones o las de sus colegas. También ha habido colaboraciones anónimas.

## Referencias

- Adler, B.T., Chatterjee, K., de Alfaro, L., Faella, M., Pye, I. and Raman, V. (2008). Assigning Trust to Wikipedia Content. *WikiSym*.
- Antin, J., Yee, R., Cheshire, C. and Nov, O. (2011). Gender differences in Wikipedia editing. *WikiSym*, pp. 11–14.
- Auer, S., Bizer, C., Lehmann, J., Kobilarov, G., Cyganiak, R. and Ives, Z. (2007). DBpedia: A nucleus for a web of open data. *ISWC*.
- Ayers, P. (2006). Researching Wikipedia – Current approaches and new directions. *Proceedings of the American Society for Information Science and Technology*.
- Ayers, P. and Priedhorsky, R. (2011). WikiLit: Collecting the Wiki and Wikipedia Literature. *WikiSym*.
- Carter, J. (2008). ClueBot and Vandalism in Wikipedia.
- Chen, C.C. and Roth, C. (2012). Citation needed: The dynamics of referencing in Wikipedia. *WikiSym*.
- Chesney, T. (2006). An empirical exploration of Wikipedia’s credibility. *First Monday*, 11(11).
- Ciglan, M. and Nørnvåg, K. (2010). WikiPop - Personalized Event Detection System Based on Wikipedia Page View Statistics.
- Collier, B. and Bear, J. (2012). Conflict, criticism, or confidence: an empirical examination of the gender gap in wikipedia contributions. *Computer-Supported Cooperative Work*, pp. 383–392.
- Danielle Elder, R.N.W. and Reilly, M. (2012). Wikipedia Lover, Not a Hater: Harnessing Wikipedia to Increase the Discoverability of Library Resources. *Journal of Web Librarianship*, 6(1):32–44.
- Encyclopædia Britannica, I. (2006). Fatally Flawed: Refuting the recent study on encyclopedic accuracy by the journal Nature.
- Faulkner, R., Walling, S. and Pinchuk, M. (2012). Etiquette in Wikipedia: Weening New Editors into Productive Ones. *WikiSym*.
- Fernanda B. Viégas, M.W. and Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. *CHI*.

- Ferretti, E., Fusilier, D.H., Cabrera, R.G., y Gómez, M.M., Errecalde, M. and Rosso, P. (2012). On the Use of PU Learning for Quality Flaw Prediction in Wikipedia. *PAN*.
- Ferron, M. and Massa, P. (2012). Psychological processes underlying Wikipedia representations of natural and manmade disasters. *WikiSym*.
- Ferschke, O., Gurevych, I. and Rittberger, M. (2012). FlawFinder: A Modular System for Predicting Quality Flaws in Wikipedia. *PAN*.
- Ford, H. and Geiger, R.S. (2012). Writing up rather than writing down: Becoming Wikipedia Literate. *WikiSym*.
- Giles, J. (2005). Special Report Internet encyclopaedias go head to head. *Nature*.
- Jullien, N. (2012). What We Know About Wikipedia: A Review of the Literature Analyzing the Project(s).
- Kittur, A. and Kraut, R.E. (2010). Beyond Wikipedia: Coordination and Conflict in Online Production Groups. *Computer-Supported Cooperative Work*.
- Krötzsch, M., Vrandečić, D. and Völkel, M. (2006). Semantic MediaWiki. *ISWC*, 4273:935–942.
- Lally, A.M. and Dunford, C.E. (2007). Using Wikipedia to Extend Digital Collections. *D-Lib Magazine*, 13(5/6).
- Leuf, B. and Cunningham, W. (2001). *The Wiki Way: Quick Collaboration on the Web*.
- Luyt, B. and Tan, D. (2010). Improving Wikipedia’s credibility: References and citations in a sample of history articles. *J. Am. Soc. Inf. Sci. Technol.*, 61(4):715–722.
- Martin, O.S. (2011). A Wikipedia Literature Review.
- Mashhadi, A., Quattrone, G., Capra, L. and Mooney, P. (2012). On the Accuracy of Urban Crowd-Sourcing for Maintaining Large-Scale Geospatial Databases. *WikiSym*.
- Massa, P., Napolitano, M. and Scrinzi, F. (2011). WikiTrip: animated visualization over time of gender and geo-location of Wikipedians who edited a page.

- Massa, P. and Scrinzi, F. (2012). Manypedia: Comparing Language Points of View of Wikipedia Communities. *WikiSym*.
- Márton Mestyán, T.Y. and Kertész, J. (2012). Early Prediction of Movie Box Office Success based on Wikipedia Activity Big Data.
- Nature (2006). Nature's responses to Encyclopaedia Britannica.
- Årup Nielsen, F. (2007). Scientific citations in Wikipedia. *First Monday*, 12(8).
- Nielsen, F.A. (2011). Wikipedia research and tools: Review and comments.
- Okoli, C. (2009). A Brief Review of Studies of Wikipedia in Peer-Reviewed Journals. *Third International Conference on Digital Society*.
- Okoli, C., Mehdi, M., Mesgari, M., Nielsen, F.A. and Lanamaki, A. (2012). The people's encyclopedia under the gaze of the sages: a systematic review of scholarly research on Wikipedia.
- Okoli, C. and Schabram, K. (2009). Protocol for a systematic literature review of research on the Wikipedia. *International Conference on Management of Emergent Digital EcoSystems*.
- Osborne, M., Petrović, S., McCreadie, R., Macdonald, C. and Ounis, I. (2012). Bieber no more: First Story Detection using Twitter and Wikipedia.
- Palomo-Duarte, M., Dodero-Beardo, J.M., Medina-Bulo, I., Rodríguez-Posada, E.J. and Ruiz-Rube, I. (2012a). Assessment of collaborative learning experiences by graphical analysis of wiki contributions. *Interactive Learning Environments*.
- Palomo-Duarte, M., Medina-Bulo, I., Rodríguez-Posada, E.J. and Palomo-Lozano, F. (2012b). Wikis en docencia: una experiencia con WikiHaskell y StatMediaWiki. *Revista de Universidad y Sociedad del Conocimiento*, 9(1).
- Palomo-Duarte, M., Rodríguez-Posada, E.J., Medina-Bulo, I. and Sales-Montes, N. (2012c). Tecnologías wiki en la docencia de Ingeniería Informática. *ReVisión*, 5(1).
- Popitsch, N., Mosser, R. and Philipp, W. (2010). Urobe: a prototype for wiki preservation. *7th International Conference on Preservation of Digital Objects*.

- Rad, H.S. and Barbosa, D. (2012). Identifying controversial articles in Wikipedia: A comparative study. *WikiSym*.
- Rector, L.H. (2008). Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review*, 36(1):7–22.
- Rodríguez-Posada, E.J. (2009). Detector y corrector automático de ediciones maliciosas en Wikipedia.
- Rodríguez-Posada, E.J. (2010). AVBOT: detección y corrección de vandalismos en Wikipedia. *NovATICA*, (203):51–53.
- Rodríguez-Posada, E.J., Dodero-Beardo, J.M., Palomo-Duarte, M. and Medina-Bulo, I. (2011). Learning-Oriented Assessment of Wiki Contributions: How to Assess Wiki Contributions in a Higher Education Learning Setting. *International Conference on Computer Supported Education*.
- Rodríguez-Posada, E.J., Ángel González Berdasco, Canduela, J.A.S., Sanz, S.N. and Saorín-Pérez, T. (2012). Wiki Loves Monuments 2011: the experience in Spain and reflections regarding the diffusion of cultural heritage. *Digithum*, (14).
- Saorín-Pérez, T. and Rodríguez-Posada, E.J. (July 2012). Análisis de enlaces hacia Bibliotecas y Archivos Digitales de Patrimonio Cultural desde Wikipedia en español y catalán. *BiD: textos universitaris de biblioteconomia i documentació*, (28).
- Sateli, B. and Witte, R. (2012). Natural Language Processing for MediaWiki: The Semantic Assistants Approach. *WikiSym*.
- Suzuki, Y. and Yoshikawa, M. (2012). Mutual Evaluation of Editors and Texts for Assessing Quality of Wikipedia Articles. *WikiSym*.
- Voss, J. (2005). Measuring Wikipedia. *International Conference of the International Society for Scientometrics and Informetrics*.
- Warncke-Wang, M., Uduwage, A., Dong, Z. and Riedl, J. (2012). In Search of the Ur-Wikipedia: Universality, Similarity, and Translation in the Wikipedia Inter-Language Link Network. *WikiSym*.
- West, A.G., Kannan, S. and Lee, I. (2010). Detecting Wikipedia Vandalism via Spatio-Temporal Analysis of Revision Metadata. *EUROSEC*, pp. 22–28.

- West, A.G. and Lee, I. (2012). Towards Content-driven Reputation for Collaborative Code Repositories. *WikiSym*.
- Wu, G., Harrigan, M. and Cuningham, P. (2012). Classifying Wikipedia Articles Using Network Motif Counts and Ratios. *WikiSym*.
- Zhang, D., Prior, K. and Levene, M. (2012). How Long Do Wikipedia Editors Keep Active? *WikiSym*.

## **Licencia**

Esta obra está bajo licencia Creative Commons Reconocimiento-CompartirIgual 3.0 Unported.