

WikiPapers, una recopilación colaborativa de literatura sobre wikis usando MediaWiki y su extensión semántica

Emilio José Rodríguez Posada⁽¹⁾, Juan Manuel Doderó Beardo⁽¹⁾

⁽¹⁾ {emiliojose.rodriguez, juanma.dodero}@uca.es
Departamento de Lenguajes y Sistemas Informáticos, C/ Chile 1, Cádiz

Resumen

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. En este artículo presentamos WikiPapers, un proyecto colaborativo para recopilar toda la literatura sobre wikis. Hasta noviembre de 2012 se han recopilado más de 1.700 publicaciones y sus metadatos, además de documentación sobre herramientas y *datasets* relacionados.

1. Introducción

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 y no cubrían todas las expectativas de los investigadores de este campo [1]. 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 artículo 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 artículo se divide de la siguiente manera. En la sección 2 hacemos 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 presentamos WikiPapers, cómo funciona y qué pasos se están dando. Finalmente, en la sección 4, terminamos con unas conclusiones y trabajo futuro.

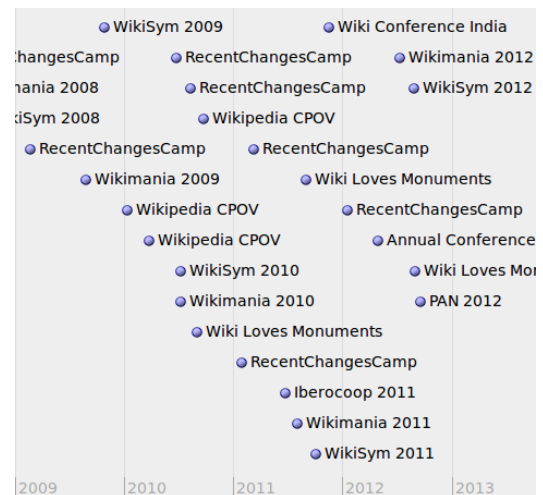


Figura 1. Línea temporal de eventos sobre wikis (mostrados solo desde 2009).

2. Trabajos relacionados

Ha habido distintos intentos de recopilar toda la literatura sobre wikis. Se han hecho recopilaciones en páginas personales y blogs, en grupos de servicios como Zotero y a través de

revisiones de literatura. A continuación los describimos y evaluamos sus ventajas e inconvenientes y cómo WikiPapers resuelve las carencias de estos enfoques.

Páginas personales y blogs: Algunos autores han hecho recopilaciones de literatura en webs personales [2] y blogs. Un ejemplo bastante completo de este último tipo es SWEETpedia [3] 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.

Gestores bibliográficos: Se han empleado gestores bibliográficos como WIKINDEX [4] creando portales propios como Wikibibliographie ENCYLEN [5], grupos en Zotero [6], BibSonomy [7], CiteULike [8] y Mendeley [9] y otros ya desaparecidos. La capacidad de estos servicios para aprovechar los metadatos generando tablas o gráficos es inexistente, pero WikiPapers las explota con sus extensiones gráficas.

Buscadores: Google Scholar, con buena cobertura, carece de representaciones visuales y no permite comentarios. Microsoft Academic Search 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.

Revisiones de literatura: Finalmente, se han realizado varias revisiones de literatura con diferente grado de exhaustividad. La primera de ellas [10] se hizo en un momento en el que las publicaciones eran escasas, pero ya se podía ver una tendencia de publicación creciente. Un año más tarde [11] hace 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 [12] 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 [13] analiza el estado del arte. Dos años más tarde [14], hace la mayor revisión de literatura en un documento de más de 50 páginas, en progreso e

inacabado, que incluye 300 referencias a publicaciones.

[15] 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.

[16] 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 [17] hace un repaso por las motivaciones para contribuir, roles, estructura, la vida de un artículo, calidad, experiencia de usuario y accesibilidad, entre otros.

El inconveniente más acuciante 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. WikiPapers

WikiPapers [18] 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. 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). Los metadatos son exportables en los formatos BibTeX, RDF, CSV y JSON. 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.

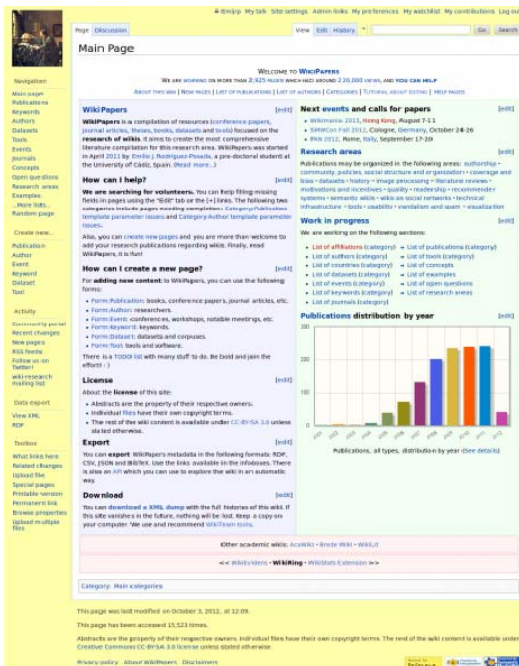


Figura 2. Portada de WikiPapers.

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

Publicaciones: En WikiPapers cada publicación dispone de una página en la que se detallan todos sus metadatos (título, autores, palabras clave, año, revista o 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. Existen listados según el tipo de publicación (artículo de revista, artículo de conferencia, tesis), por idioma, año, revista, conferencia, licencia, número de citas, etc.

Palabras clave: Existe un listado de todas las palabras clave 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 wiki, web 2.0, collaboration, evaluation, collaborative learning, knowledge management, MediaWiki, motivation, data mining y conflict.

Tabla 1. Requisitos sugeridos en WikiSym

| 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 |

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).

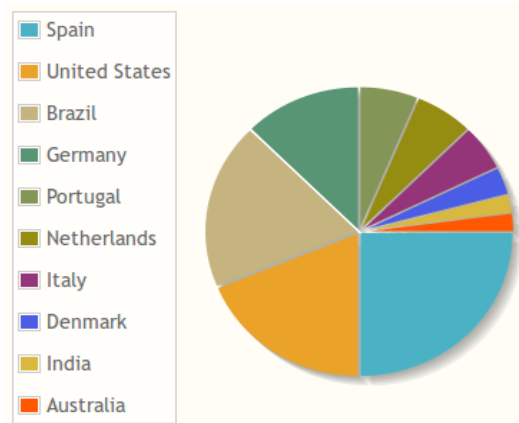


Figura 3. Autores para los que se conoce su nacionalidad.

Datasets: Hay un listado de *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 [19] está compilando una gran cantidad de datos sobre comunidades wiki, que se cifra ya en 4.500 *dumps*.

Herramientas: Se está construyendo un listado de herramientas que se han desarrollado a la hora de investigar sobre wikis y para proponer soluciones a ciertos problemas. Entre ellas se incluyen herramientas sobre anti-vandalismo, estadística y visualización, *frameworks*, lenguaje, preservación y procesamiento de datos.

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.

4. Conclusiones y trabajo futuro

El estudio de los wikis es un campo emergente y prolífico. Hasta ahora la recopilación de toda la literatura sobre wikis se había realizado en webs personales, blogs, gestores bibliográficos, servicios web, redes sociales, pero cada uno tiene sus ventajas e inconvenientes. Ninguno de ellos ha conseguido ser exhaustivo, permanecer actualizado con el paso del tiempo y satisfacer a los investigadores del campo de los wikis. En este artículo hemos presentado WikiPapers, un proyecto colaborativo para recopilar toda la literatura sobre wikis. Se hace uso de MediaWiki y su extensión semántica.

En noviembre de 2012 cuenta ya con información de más de 1.700 publicaciones, 1000 autores, 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 descarga y facilitar su preservación. WikiPapers es una comunidad abierta a la participación. Estamos recopilando colaborativamente toda la literatura sobre wikis... ¡en un wiki!

5. Agradecimientos

Agradecimientos para los colaboradores voluntarios de WikiPapers que han ayudado a añadir publicaciones y a mejorar los metadatos de las existentes. También para Izaskun Mallona

que ha revisado este artículo y ayudado con sus consejos.

6. Referencias

- [1] P. Ayers y R. Priedhorsky, WikiLit: Collecting the Wiki and Wikipedia Literature. WikiSym (2011). <http://etherpad.wikimedia.org/WikiLit>
- [2] <http://www.public.iastate.edu/~CYBERSTACS/WikiBib.htm>
- [3] <http://www.mkbergman.com/sweetpedia/>
- [4] <http://sourceforge.net/projects/wikindx/>
- [5] http://wikindx.inrp.fr/biblio_encyclen/
- [6] http://www.zotero.org/groups/wikipedia_research
- [7] <http://www.bibsonomy.org/tag/wiki>
- [8] <http://www.citeulike.org/tag/wiki>
- [9] <http://www.mendeley.com>
- [10] J. Voss, Measuring Wikipedia, *International Conference of the International Society for Scientometrics and Informetrics* (2005).
- [11] P. Ayers, Researching Wikipedia - Current approaches and new directions, *Proceedings of the American Society for Information Science and Technology* (2006).
- [12] C. Okoli y K. Schabram, *Protocol for a systematic literature review of research on the Wikipedia*, International Conference on Management of Emergent Digital EcoSystems (2009).
- [13] C. Okoli, A Brief Review of Studies of Wikipedia in Peer-Reviewed Journals (2009).
- [14] F. A. Nielsen, Wikipedia research and tools: Review and comments (2011).
- [15] S. O. Martin, A Wikipedia Literature Review, *arXiv* (2011).
- [16] C. Okoli, M. Mehdi, M. Mesgari, F. A. Nielsen, A. Lanamaki, The people's encyclopedia under the gaze of the sages: a systematic review of scholarly research on Wikipedia, *Social Science Research Network* (2012).
- [17] N. Jullien, What We Know About Wikipedia: A Review of the Literature Analyzing the Project(s), *Social Science Research Network* (2012).
- [18] <http://wikipapers.referata.com>
- [19] <http://code.google.com/p/wikiteam/>