Influencias Geoculturales en el Uso de Wikipedia: mejorando el acceso, la diversidad y la relevancia en los buscadores

Javier López Otero
Universidad Isabel I

Ángel Obregón Sierra
Universidad Isabel I

Antonio Gavira Narváez
Universidad Isabel I

Resumen

Este proyecto analiza cómo los entornos geográficos y culturales influyen en las búsquedas en Wikipedia para mejorar la accesibilidad y representatividad del contenido. Ello implicará el análisis de patrones de búsqueda y la influencia cultural y lingüística, utilizando Google Trends y GIS. Se recopilarán datos mediante encuestas, grupos focales, minería de datos y se analizarán mediante machine learning. Los resultados incluirán un mapa interactivo de consultas, informes sobre diversidad cultural, modelos de análisis predictivo, guías de optimización de contenido, estrategias de expansión, modelos de participación comunitaria y un análisis de impacto en el acceso al conocimiento, beneficiando a editores, investigadores y a Wikimedia.

Introducción

En la era digital, Wikipedia ha pasado a ser un recurso clave de conocimiento accesible y gratuito, ampliando su alcance a través de dispositivos móviles. Sin embargo, hay una falta de comprensión sobre cómo el entorno geográfico y cultural de los usuarios influye en sus consultas en Wikipedia. Así pues, este proyecto pretende analizar cómo dichos entornos específicos influyen en los patrones de búsqueda en Wikipedia. Esta información se usará para mejorar la experiencia del usuario, al favorecer el acceso a información mejor adaptada a la consulta de intención del usuario. Este análisis puede ser esencial para la Fundación Wikimedia, ya que la comprensión de las dinámicas es esencial para mantener y mejorar la relevancia y eficiencia de Wikipedia.

Fecha: 1 de junio de 2024 hasta el 1 de junio de 2025.

Trabajos relacionados

Existen estudios relacionados, por ejemplo, Miquel-Ribé y Laniado (2018, 2019), que analizan la diversidad cultural en Wikipedia, y Teplitskiy, Lu y Duede (2017) que analizaron la influencia del estatus académico y la accesibilidad en las referencias de Wikipedia. Otros trabajos afines son el de Meeks (2010) sobre cartografía en Wikipedia, Hu et al. (2009) en intención de consulta, Kousha y Thelwall (2016) sobre el impacto académico de las citas de Wikipedia, y Lewandowski y Spree (2011) sobre la calidad de los artículos en buscadores. Sin embargo, aunque existen trabajos afines el impacto directo del contexto geográfico-cultural en las consultas de Wikipedia está poco estudiado.

Método

Fases de la metodología:
1.- Recolección de datos:
Análisis de consultas de usuarios en buscadores: Se emplearán herramientas como Google BigQuery, Google Trends y Google Public Data Explorer para extraer datos de consultas de usuarios, enfocándonos en la ubicación y el idioma de estas búsquedas.
1.1. Datos espaciales: Mapeo de las consultas espaciales con GIS, apoyandonos en Google Maps y Bing Maps. El área geográfica de estudio será en 5 ámbitos y en 5 países o regiones con idiomas diferentes.
1.2. Datos no espaciales: Analizaremos la posición de Wikipedia en los resultados de búsquedas no espaciales en distintos idiomas.
1.3. Encuestas y grupos focales:
Diseño de Encuestas: encuestas para usuarios activos y no activos en Wikipedia, estudiando sus hábitos y preferencias de búsqueda.
Se organizarán grupos focales para entender las preferencias y comportamientos de búsqueda.
2. Análisis de Datos:

2.1.Análisis de patrones en consultas: minería de datos y análisis estadístico para examinar patrones en las consultas espaciales y temporales, considerando variables como la ubicación geográfica, el idioma y la temática.
Desarrollo de modelos predictivos machine learning para modelar las relaciones entre ubicación, contexto cultural y lingüístico, y consultas. Los modelos integrarán datos de consultas, percepciones de usuarios y métricas internas de artículos de Wikipedia.

Rendimiento esperado

Elaboración de un mapa interactivo con las tendencias de uso de Wikipedia. Identificará las áreas geográficas y lingüísticas que requieren más atención.

Elaboración de modelos de análisis predictivo para ayudar a los editores a anticipar y responder a necesidades de información emergentes.

Las guías de optimización de contenido según tendencias de búsqueda aumentarán la visibilidad y accesibilidad del contenido.

El público objetivo de esta propuesta son editores, administradores y usuarios de Wikipedia y académicos.

Riesgo

Puede ser un reto conseguir participantes representativos para los grupos focales.
Las mejoras implementadas en Wikipedia pueden encontrar resistencia o desafíos prácticos, limitando el impacto del proyecto.

Plan de impacto

La maximización del impacto del proyecto será coherente con la Estrategia del Movimiento Wikimedia 2030. Asimismo, se realizarán acciones como: colaborar con comunidades de Wikimedia, participar en eventos de Wikimedia, publicar resultados en formatos académicos y divulgativos.

Evaluación

La evaluación del impacto del proyecto en las consultas a Wikipedia según entornos geográficos y culturales requiere criterios de éxito claros y medibles. Incluye evaluar la eficiencia en recolección y análisis de datos para identificar patrones de búsqueda, la efectividad de encuestas y grupos focales en términos de calidad y representatividad de datos, la precisión y robustez de modelos predictivos, y la relevancia y utilidad de los resultados para la comunidad de Wikimedia.

Presupuesto

Salario para tres técnicos a tiempo parcial especialistas en análisis de datos estadísticos, espaciales e informáticos. 30.000 €
Costos de publicación en acceso abierto. 4.000 €
Gastos generales institucionales. 3.400€
Total solicitado. 37.400 €.

Principales contribuciones

Ángel Obregón es bibliotecario en Wikipedia en español (usuario Vanbasten_23). Lleva 17 años colaborando en los proyectos de la Fundación Wikimedia, especialmente en apartados técnicos, programando bots en varios idiomas y Wikidata. Ha publicado 14 artículos científicos sobre Wikipedia, principalmente en educación, y 4 sobre Wikidata.

Javier López es profesor de la Universidad Isabel I, ha publicado más de 20 artículos y capítulos de libro, ha participado en 5 proyectos de I+D+I en España y ha presentado más de 10 ponencias en el extranjero.

Antonio Gavira es profesor de la Universidad Isabel I, ha publicado más de 30 artículos, capítulos de libro y comunicaciones en congresos. También, ha participado en varios proyectos de I+D+I de temática tecnológica.

La experiencia investigadora más reciente en Wikipedia sería : *Methodology for the Incorporation of Geographic Information in Wikidata* y *Unveiling Wikipedia's Role in Urban Tourism: An In-depth Analysis of Destination Choice Using Artificial Intelligence Indicators* (en evaluación).

Anexo 1: Metodología

Referencias

Hu, J., Wang, G., Lochovsky, F., Sun, J. T., & Chen, Z. (2009). Understanding user's query intent with wikipedia. En *Proceedings of the 18th international conference on World wide web,* 471-480. https://doi.org/10.1145/1526709.1526773

Kousha, K., & Thelwall, M. (2016). Are Wikipedia citations important evidence of the impact of scholarly articles and books?. *Journal of the Association for Information Science and Technology*, *68*(3), 762-779. https://doi.org/10.1002/asi.23694

Lewandowski, D., & Spree, U. (2011). Ranking of Wikipedia articles in search engines revisited: Fair ranking for reasonable quality?. J. Am. Soc. Inf. Sci., 62, 117-132. https://doi.org/10.1002/asi.21423

Lewoniewski, W., Härting, R.C., Węcel, K., Reichstein, C., & Abramowicz, W. (2018). Correction to: Application of SEO Metrics to Determine the Quality of Wikipedia Articles and Their Sources. En: Damaševičius, R., Vasiljevienė, G. (eds) *Information and Software Technologies*. ICIST 2018. Communications in Computer and Information Science, 920. Springer, Cham. https://doi.org/10.1007/978-3-319-99972-2_49

López-Otero, J., Gaviran-Narvaez, A., & Vega-Pozuelo, R.F. (2024). *In-depth Analysis of Destination Choice Using Artificial Intelligence Indicator*s. Social Indicators Research (en evaluación)

Meeks, E. (2011). Mapping Wikipedia: Geolocated Articles as a Proxy of Culture and Attention Stanford University. https://dhs.stanford.edu/spatial-humanities/mapping-wikipedia-geolocated-articles-as-a-proxy-of-culture-and-attention

Miquel-Ribé, M., & Laniado, D. (2018). Wikipedia Culture Gap: Quantifying Content Imbalances Across 40 Language Editions. *Front. Phys. 6*(54). https://doi.org/10.3389/fphy.2018.00054

Miquel-Ribé, M., & Laniado, D. (2019). Wikipedia Cultural Diversity Dataset: A Complete Cartography for 300 Language Editions. *Proceedings of the International AAAI Conference on Web and Social Media, 13*(1), 620-629. https://doi.org/10.1609/icwsm.v13i01.3260

Obregón-Sierra, A., López-Otero, J. Gavira-Narváez, A., & Vega-Pozuelo, R.F. (2023). Methodology for the Incorporation of Geographic Information in Wikidata. Revista De Estudios Andaluces, 45, 260-64. https://revistascientificas.us.es/index.php/REA/article/view/22878/20653

Teplitskiy, M., Lu, G., & Duede, E. (2017). Amplifying the impact of open access: Wikipedia and the diffusion of science. *Journal of the Association for Information Science and Technology*, *68*(9), 2116-2127. https://doi.org/10.1002/asi.23687