

Construcción del conocimiento alrededor de sistemas híbridos fotovoltaicos mediante Descubrimiento basado en la literatura

Knowledge construction around hybrid photovoltaic systems through Literature Based Discovery

Marco Coral ^{1,a}, Miguel Salazar ^{1,b}, Luis Palacios ^{1,c}, Santiago Moquillaza ^{1,d}, José C. Álvarez ^{2,e}

¹ Universidad Nacional Mayor de San Marcos. Lima, Perú

² Universidad Peruana de Ciencias Aplicadas. Lima, Perú

^a mcoraly@unmsm.edu.pe

^b miguel.salazarkovaleff@unmsm.edu.pe

^c luisalberto.palacios@unmsm.edu.pe

^d smoquillazah@unmsm.edu.pe

^e pcijjalv@upc.edu.pe

Resumen

Ante la necesidad de profundizar la dinámica en la generación de nuevos conocimientos, con relación a las energías renovables, en particular alrededor de los sistemas híbridos fotovoltaicos, en este estudio se examinan las investigaciones elaboradas bajo una estructura de colaboración de investigación, mediante "Literature Based Discovery" - LBD. Se elabora una cadena de búsqueda, para capturar aspectos emergentes y otros aspectos innovadores o exploratorios de energía renovable de los sistemas híbridos fotovoltaicos. Los pasos seguidos tienen como prioridad generar la extracción y búsqueda automatizada, a partir de las palabras-clave, agrupar los resultados buscando las referencias de estos, además de ubicar investigaciones sobre estos términos semánticos, conceptos y la relación entre ellos mediante minería de texto. Los resultados generan agrupación de tópicos en base a las palabras clave, los que han sido vinculados por LBD, estos identifican dependencias entre diversos análisis concluyéndose que las técnicas LBD generan correlaciones que no se ven explícitamente, por lo que es necesario realizar análisis adicionales por parte del experto para determinar correctamente el conocimiento implícito.

Palabras clave: Literature Based Discovery; sistemas híbridos fotovoltaicos; agrupación; minería de texto; conocimiento implícito.

Abstract

Given the need to deepen the dynamics in the generation of new knowledge, with respect to renewable energies, in particular around hybrid photovoltaic systems, this study analyzes the research developed, through the "Literature Based Discovery" - LBD. A research chain is developed to capture emerging and other innovative or exploratory aspects. The steps followed have as a priority to generate the extraction and automated search, from the key- words, to group the results looking for their references, in addition to locating research on these semantic terms, concepts and the relationship between them through text mining. The results generate a grouping of topics based on the keywords, which were linked by LBD, these identify dependencies between various analyzes, concluding that LBD techniques generate correlations that are not explicitly seen, so it is necessary to carry out additional analyzes by the expert to correctly determine the implicit knowledge.

Keywords: Literature Based Discovery; hybrid photovoltaic systems; clustering; text mining; implicit knowledge.

Correspondencia:

Dirección: Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Lima, Perú.

Recibido 20/10/2020 - Aceptado 28/11/2020 - Publicado: 21/12/2020

Citar como:

Coral, M., Salazar, M., Palacios, L., Moquillaza, S. & Álvarez, J. (2020) Construcción del conocimiento alrededor de sistemas híbridos fotovoltaicos mediante Descubrimiento basado en la literatura. Revista Peruana de Computación y Sistemas, 3(2):11-18. <http://dx.doi.org/10.15381/rpcs.v3i2.19257>

© Los autores. Este artículo es publicado por la Revista Peruana de Computación y Sistemas de la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional Mayor de San Marcos. Este es un artículo de acceso abierto, distribuido bajo los términos de la licencia Creative Commons Atribución - No Comercia_Compartir Igual 4.0 Internacional. (<http://creativecommons.org/licenses/by-nc-sa/4.0/>) que permite el uso no comercial, distribución y reproducción en cualquier medio, siempre que la obra original sea debidamente citada.

1. Introducción

Los sistemas híbridos fotovoltaicos pueden describirse como la combinación de tecnologías de energía solar con otras fuentes de energía para la generación eléctrica, con el objetivo de dar respuesta inmediata al consumo (demanda eléctrica), se tienen sistemas integrados y sistemas fuera de línea (off grid). El término off-the-grid, señala a una aplicación fuera de la red, es decir se refiere a un sistema sin red (aislada o autónoma), que inicialmente no está conectado a una red eléctrica principal o nacional, se toma este tema como base para nuestra investigación debido a su importancia en el campo energético y su relevancia en la sociedad además de existir bastantes publicaciones en diferentes Bases de Datos.

La generación de conocimiento a través de literature based discovery (LBD), es importante porque permite identificar el conocimiento no explícito mediante ciertas técnicas computacionales, permite construir asociaciones de búsqueda potencial, mediante el análisis de la literatura, logrando con ello la construcción acelerada de conocimiento y de manera heterodoxa, que permite ingresar en un nuevo tema de investigación actualizado, a través de la generación de hipótesis en el marco de investigación [1]. A través de LBD, se establece la forma convencional donde se quiere hacer la extrapolación de un tópico "A" conectado a otro "C", a través de un tópico intermedio "B", para esto se correlaciona el conocimiento utilizando formas automatizadas para extraer y analizar tendencias estadísticas entre ellos podemos citar el análisis de Cluster y el análisis de frecuencias [2], [3] y [4], con lo cual se pueden generar diagramas de influencia, y diagramas de densidad.

El propósito de esta investigación es identificar la construcción del conocimiento alrededor de sistemas híbridos fotovoltaicos mediante LBD, para ello utilizamos clustering y text-mining, por ello accedemos al grupo selecto de los hallazgos, de las fuentes potenciales de información académica y científica en la Base de Datos Scopus, tomamos las relaciones generadas a partir de las palabras claves (Key-words) de cada artículo científico para iniciar nuestro análisis.

2. Estado del arte

La construcción del conocimiento mediante LBD, hace referencia a la utilización de técnicas computacionales que permiten la generación automática del mismo, algunos autores buscan construir modelos computacionales a fin de interpretar el conocimiento implícito [5], a partir de análisis de frecuencias, conocimiento parcial, aprendizaje basado en reglas, modelos semánticos, análisis de redes, entre otros. Entre las técnicas más utilizadas en LBD [3] y [6], tenemos; Bibliometrics Analysis [7], Clustering, Network analysis, text-mining, Query enhancements, Distributional Semantic, Structured Knowledge, Bases-Ontologies-Taxonomies, etc., en los últimos años text-mining ha surgido como una alternativa importante en la extracción

de conocimiento basado en la semántica y el análisis de textos [8], por tanto es útil para concentrarnos en la búsqueda de investigaciones de energías renovables en la modalidad de sistemas híbridos fotovoltaicos (sistemas aislados). Moro [9] sostiene que el análisis de text-mining puede realizarse a partir del análisis de palabras claves para ello utiliza el software "Tools for Innovation Monitoring" (TIM), el cual extrae y analiza las palabras clave relevantes de diversas publicaciones científicas con el objetivo de extraer conocimiento para los diversos sistemas energéticos, Trappey [10] por su parte utiliza text-mining para categorizar datos asociados a la energía solar elaborando un mapa de relaciones basado en palabras claves, luego aplica jerarquías analíticas para evaluar la meta usando un algoritmo de clusterización para segmentar las características de las patentes.

Según Porter y Zhang [11], la base del proceso de extracción es un diccionario de text-mining personalizado desarrollado por el Centro Común de Investigación (JRC) de la Comisión Europea, y optimizado para texto científico, en donde los conceptos del diccionario pueden contener muchas versiones de la misma palabra (célula solar, células solares, célula solar ...), por lo que palabras similares son agrupados en conceptos de acuerdo con el proceso conocido, en la literatura, como "agrupamiento". Por su parte Vicente-Gomila [12] resalta los aportes de Semantic-Syntactic TRIZ a la técnica LBD, a través del enfoque de sistemas y subsistemas, donde el componente "A" estaría delimitado por el problema a resolver, el componente "B" por las causas y la influencia de otros subsistemas, en tanto que "C" corresponde con la solución al problema.

3. Metodología

Esta investigación sigue un método de revisión sistemática de la literatura en el tema en cuestión [1], se adopta el enfoque cualitativo basado en un análisis bibliométrico con el cual se busca conocer la actividad científica relevante en función a los temas relacionados con Sistemas Fotovoltaicos, también sigue un enfoque cualitativo para realizar el análisis de contenido de la literatura especializada. El primer paso de la revisión bibliográfica se refiere a la recopilación de literatura relevante sobre el tema, para este propósito, analizamos las bases de datos de Scopus. La Tabla 1 siguiente resume la metodología de investigación. La búsqueda de datos se realizó el 08 de julio del 2020.

Tabla 1. Metodología de la Investigación

Enfoque	Descripción	Contenido
Enfoque cuantitativo	Una caracterización cuantitativa de las publicaciones seleccionadas.	Cantidad de publicaciones. Cantidad de publicaciones segmentadas por tema y palabras clave.
Enfoque cualitativo	Análisis de contenido de artículos basado en LBD.	Análisis de frecuencias, Agrupaciones, proyecciones, análisis de textos.

Para las Bases de Datos consultadas, los parámetros de búsqueda se definieron según se muestra en la tabla 2 los resultados obtenidos fueron 29 artículos en SCOPUS, de los cuales se filtraron las palabras claves utilizadas en la indexación de los artículos.

Tabla 2. Revisión sistemática de la Literatura

Criterio	Filtros	artículos
Palabras claves	TITLE-ABS-KEY ("renewable energies" AND "photovoltaic systems" AND "off grid" AND "hybrid systems") AND (LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015)) AND (LIMIT-TO (SUBJAREA , "ENER") OR LIMIT-TO (SUBJAREA , "ENVI") OR LIMIT-TO (SUBJAREA , "ENGI"))	29
Áreas	Energy, Environmental Science, Engineering	
Tipo	Articles and conference proceedings	
Lenguaje	English	

Fuente: SCOPUS

4. Resultados

La cantidad de artículos publicados sobre temas relacionados con Sistemas Fotovoltaicos asociados a energía renovable, sistemas híbridos y sistemas fuera de línea, son relativamente pocos, se identificaron 29 trabajos asociados a esos temas en la Base de Datos de SCOPUS, de los cuales se visualiza que hay un mayor flujo de trabajos en el 2016 con 7 publicaciones y de 6 en el 2018 y 2020 (Fig. 1), vemos también que las citas realizadas a estos artículos suman 654 en total (Fig. 2), de estas cifras vemos que los últimos 4 artículos no tienen citas registradas lo cual se puede justificar por qué son del año 2020 y uno del 2019, no necesariamente este indicador descalifica el conocimiento generado en estos trabajos, vemos que el artículo 1 publicado en el 2016 tiene 115 citas, seguido de 76 y 73 citas para los artículos siguientes, estos indicadores muestran que el conocimiento generado es utilizado y aprovechado por otros autores y son relevantes dentro del área. Se aborda también un análisis temático en función al conocimiento representado en las palabras clave, se tiene 160 palabras claves utilizadas en los 29 artículos, de los cuales se realiza un conteo a partir de las Frecuencias y se muestran las 20 más utilizadas, se visualiza que Off-grid hybrid system se menciona en 10 artículos, lo cual nos da una idea de la importancia de la misma, seguido de Organic photovoltaics con 7 (Tabla 3), estos indicadores a su vez son agrupados a partir del juicio experto en función a la similitud temática o la referenciación a temas similares, la tabla 4 muestra las frecuencias de los 30 primeros grupos de 64 encontrados, de estas agrupaciones podemos determinar que el mayor conocimiento investigado por los diferentes autores se centra en Off grid system con 18 incidencias, seguido de hybrid energy system con 15 y photovoltaic system con 11, se observa que las investigaciones se centran en estos temas prioritariamente

los cuales son el eje temático para tratar u asociar otros temas, se elabora un análisis de densidad en el software estadístico R el cual se muestra en la Figura 3, en el podemos verificar una mayor concentración de grupos que tocan un tema en particular, lo cual significa que los temas principales se tratan y asocian con múltiples temas creando una gama de conocimientos particularizados.

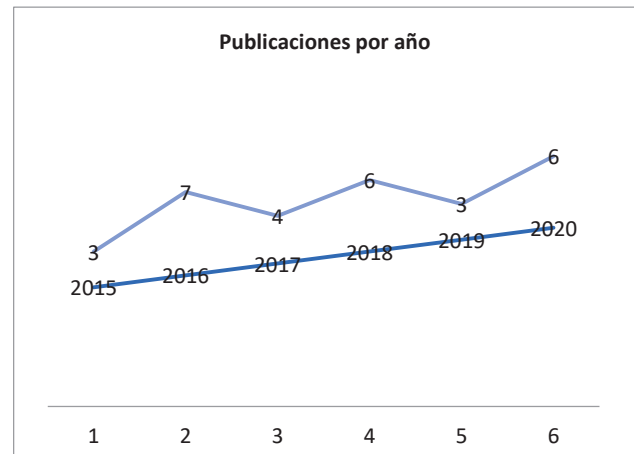


Figura 1. Artículos por Años de publicación.

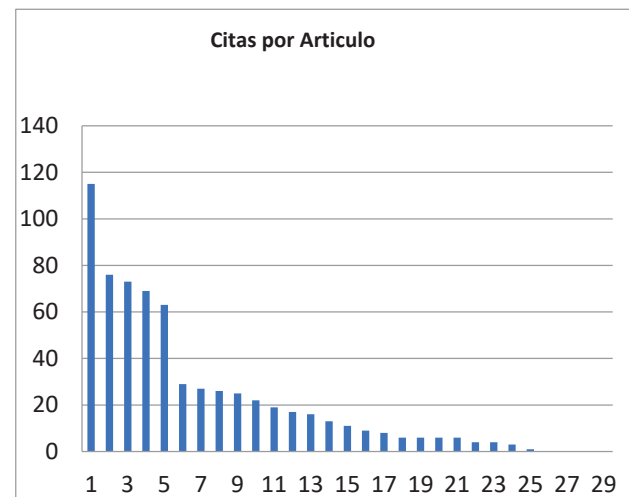


Figura 2. Citas por artículos (29)

Tabla 3. Palabras claves por frecuencia.

Palabras Claves	Frecuencia	Palabras Claves	Frecuencia
Off-grid hybrid system	10	Hybrid system	2
Organic photovoltaics	7	Life cycle assessment	2
Concentrated photovoltaics	5	Microgeneration	2
Fuel cell	4	Modelling	2
Battery storage	3	Mppt	2
Biomass gasification	3	Off-grid energy	2
Homer software	3	Photovoltaic (pv)	2
Hybrid energy system	3	Photovoltaics (pv)	2
Diesel replacement strategy	2	Renewable resources	2
Emissions	2	Shuffled frog leaping algorithm	2

Elaboración Propia

Tabla 4. Palabras claves agrupadas por frecuencia, se muestran los primeros 30 grupos. *Elaboración Propia*

Agrupaciones y Palabras claves Consideradas en el Análisis	G.	F.	Agrupaciones y Palabras claves Consideradas en el Análisis	G.	F.	Agrupaciones y Palabras claves Consideradas en el Análisis	G.	F.
Off grid system, off-grid, off-grid electrification, off-grid energy, off-grid hybrid system off-grid pv system, off-grid renewable energy system, off-grid system	1	18	Hybrid, hybrid energy system, hybrid energy systems, hybrid mini-grid, hybrid pv-diesel system, hybrid renewable, hybrid renewable energy system, hybrid solar pv/tidal, hybrid system, hybrid systems , pv hybrid systems, csp-dg hybrid system	2	15	Photovoltaic, photovoltaic (pv), photovoltaic arrays, photovoltaic system, photovoltaic systems, photovoltaic with battery, photovoltaic with diésel, photovoltaics, photovoltaics (pv)	3	11
Organic photovoltaics	4	7	Renewable energies, renewable energy, renewable energy fraction, renewable resources, renewable sources	5	6	Biomass, biomass and hybrid systems, biomass gasification	6	5
Concentrated photovoltaics	7	5	Homer, homer pro, homer software	8	5	Battery capacity degradation, battery storage	9	4
Fuel cell	10	4	Life cycle assessment, life cycle assessment (lca), life-cycle cost	11	4	Cost analysis, cost of energy, cost-effective	12	3
Diesel replacement strategy, diesel system	13	3	Emissions, emissions intensity, co2 emissions	14	4	Optimisation, optimization, optimum components	15	3
Solar photovoltaic, solar pv, solar system design program	16	3	Techno-economic analysis, techno-economical analysis, techno-economics	17	3	Simulation, matlab/simulink	22	3
Microgeneration	18	2	Modelling	19	2	Mppt	20	2
Shuffled frog leaping algorithm	21	2	Stand-alone energy system, stand-alone pv system	23	2	Tidal energy, tidal-powered desalination	24	2
Wind, wind turbines	25	2	Remote area , remote areas electrification	26	2	Bangladesh	27	1
Cameroon	28	1	Concentrating solar thermal	29	1	Desalination	30	1

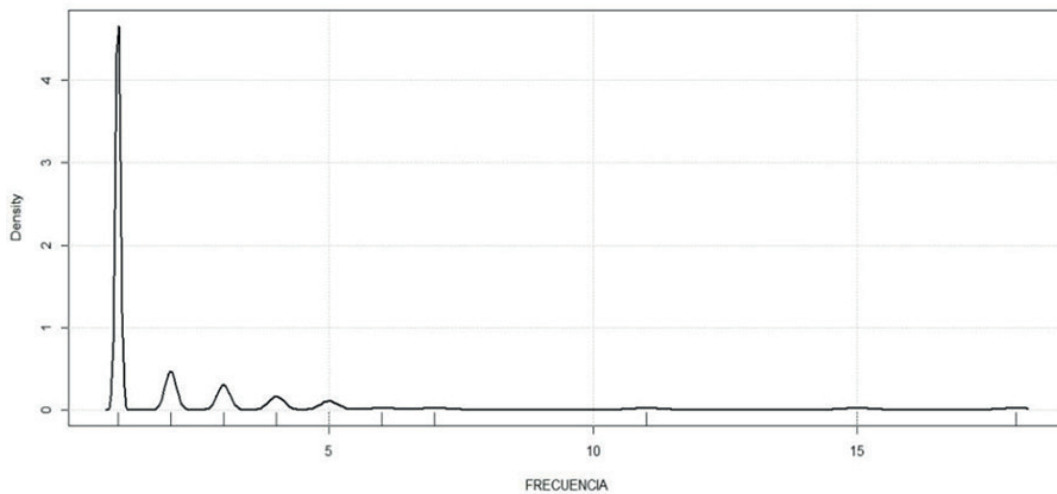


Figura 3. Análisis de densidad de frecuencias

Con respecto a los factores identificados se tiene la frecuencia y los grupos creados, con esto se busca explicar la cantidad de temas tocados por los diferentes autores, para ello se crea un diagrama de influencias utilizando R (Figura 4) teniendo que el grupo 1 y 2 son los más influyentes, la gráfica explica la cantidad de veces que fue tocado ese tema debido a su relevancia ya que se encuentra sobre un índice mayor, vemos que los siguientes grupos muestran menor influencias sobre el tema principal.

Este análisis fundamenta el hecho que los temas con frecuencias menores gasificación, fuel, greenhouse,

mitigation entre otros se combinan con los temas del grupo 1 y 2, generando relaciones donde se muestran nuevos campos o investigaciones.

Para reforzar los resultados obtenidos se utiliza text-mining bajo el software Iramuteq 0.7 alpha 2 [13], [14] y [15], para ello se parte del mismo principio es decir analizar las palabras claves e identifica los conocimientos sobre sistemas híbridos fotovoltaicos, aunque otros autores han tratado temas similares utilizando herramientas de software [9] los enfoques de los análisis varían respecto al objetivo.

El análisis realizado extrae las palabras clave relevantes del corpus de publicaciones científicas indicadas anteriormente, se preparan las variables y datos correspondientes para el análisis utilizando el sistema de codificación utf_8_sig en Inglés, la tabla 5 muestra el esquema de ingreso de datos para las 10 primeras variables correspondientes a los 10 primeros artículos, el Software reconoce 29 segmentos de texto (asociados a las palabras claves de los 29 paper), 348 ocurrencias entre formas activas y suplementarias y 83 hápax (ocurrencias únicas), el primer paso es realizar un análisis de concordancia basado en la Lematización [16], [19], el cual da como

resultado un análisis de frecuencia de las formas activas y suplementarias, la tabla 6 muestra las concordancias a nivel de texto.

A partir del análisis de concordancia se realiza el Análisis Factorial de Correspondencia (AFC) [17], [18] mostrado en la tabla 6, este análisis hace una comparativa sobre el uso de formas o unidades léxicas específicas, en este caso los textos identificados en el análisis de frecuencias que básicamente se refieren a las palabras clave que se repiten más en cada grupo, los grupos de palabras claves representan una variable por tanto la correspondencia identifica las asociaciones directas entre

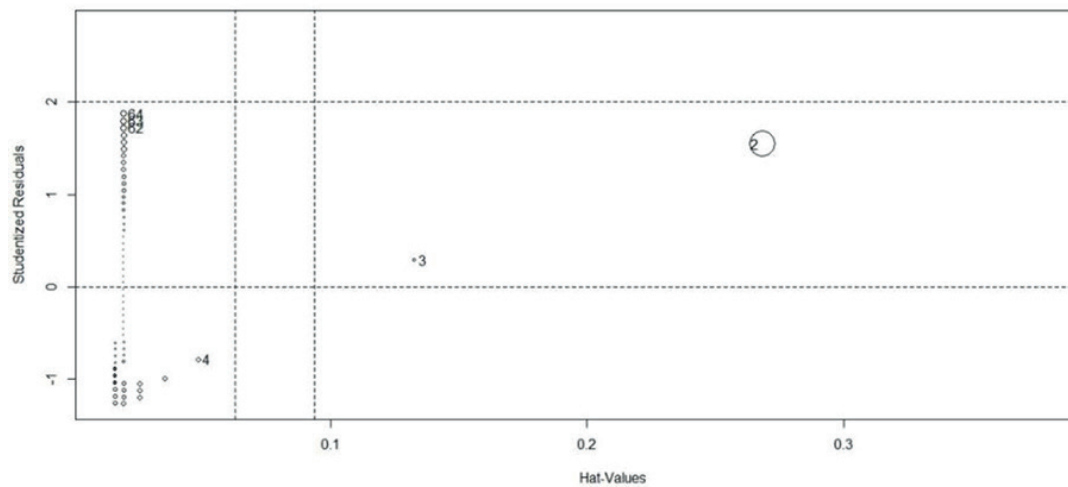


Figura 4. Diagrama de Influencia

Tabla 5. Definición de Variables y preparación de los datos

Variables	Palabras Clave
**** *K1	Diesel system; Economic analysis; HOMER; Hybrid energy systems; Photovoltaic; Wind
**** *K2	Economic distance limit; Grid-connected hybrid system; HOMER; Off-grid hybrid system; Rural electrification
**** *K3	Hybrid energy system; Photovoltaic; Renewable energy fraction; Rural electrification; Wind
**** *K4	Biomass; Cost-effective; Photovoltaic; Reliability
**** *K5	Energy dispatching; Hydrogen system; Off-grid system; Predictive control; Renewable energy
**** *K6	Anchor-Business-Community model; Biomass and hybrid systems; Off-grid renewable energy system; Solar PV
**** *K7	Hybrid energy systems; Minimal annual cost; Optimum components
**** *K8	Battery capacity degradation; LiCoO2 battery; Off-grid system; Photovoltaic system
**** *K9	Concentrated photovoltaics; Emissions intensity; Greenhouse gas mitigation; Hybrid system; India; Life cycle assessment; Off grid system; Organic photovoltaics; Photovoltaics; Rural electrification
**** *K10	HOMER; Hybrid system; Optimization; Renewable resources; Techno-economical analysis

Tabla 6. Análisis de concordancia por distribución de frecuencias. Obtenido del software Iramuteq

	*K1	*K10	*K11	*K12	*K13	*K14	*K15	*K16	*K17
System	0.4868	0.2612	-0.3449	0.3836	-0.8195	-0.3097	-0.2289	0.1839	-0.1139
Photovoltaic	0.4413	-0.1161	0.6514	-0.0771	0.9651	-0.2755	-0.0771	-0.1555	-0.0384
Hybrid	0.2016	0.3514	0.3514	0.4899	-0.1936	-0.1936	0.4899	0.2631	0.7522
Energy	0.1787	-0.2807	-0.2807	-0.1863	-0.2245	0.394	-0.1863	0.2372	-0.0928
Off	-0.1767	-0.1051	-0.1051	-0.0698	0.3596	0.3596	-0.0698	-0.1408	-0.0348
Grid	-0.2326	-0.1384	0.564	-0.0919	0.2753	0.2753	-0.0919	-0.1853	-0.0458
Renewable	-0.3291	0.4402	-0.1958	-0.13	0.1824	0.1824	0.5872	0.3437	-0.0647

cada tema en función a dichas similitudes, lográndose correspondencia positiva donde se repite el texto y correspondencia negativa que indica el uso del texto en el grupo, el análisis detallado se muestra en la gráfica de la figura 5. Las agrupaciones por color en el eje positivo muestran las coincidencias porcentuales en los 29 artículos los cuales son la base para identificar correspondencias de temas.

A partir del Análisis Factorial de Correspondencia se determinan los resultados observados en la gráfica de la figura 6, se observa que los temas tocados en el paper 4 (Biomass, Cost-effective, Photovoltaic, Reliability) son tratados muy pocas veces y son de gran relevancia, mientras que los temas tratados en el paper 17 (HOMER; Hybrid; Insolation; Optimisation; Photovoltaics)

son tratados pocas veces y son de poca relevancia, los temas tratados en el paper 20 (HOMER PRO; Hybrid; Renewable energy; Stand-alone energy system) son tratados muchas veces y tienen poca relevancia, por ultimo vemos el paper 28 (Battery storage; Biomass; CO2 emissions; Cost analysis; Feasibility analysis; Off-grid electrification; Payback period; Photovoltaics (PV); Renewable energy; Simpl) son tratados muchas veces y tienen alta relevancia, de esta grafica podemos identificar los conocimientos que tienen alta relevancia de los que no lo tienen y los que son tratados con mayor frecuencia, a partir de esta grafica podemos inferir los focos de interés en este grupo de conocimientos. Debemos indicar también que los 64 grupos definidos a partir de las similitudes de palabras claves, pueden relacionarse entre ellos formando nuevas asociaciones, la Figura 7 nos indica las

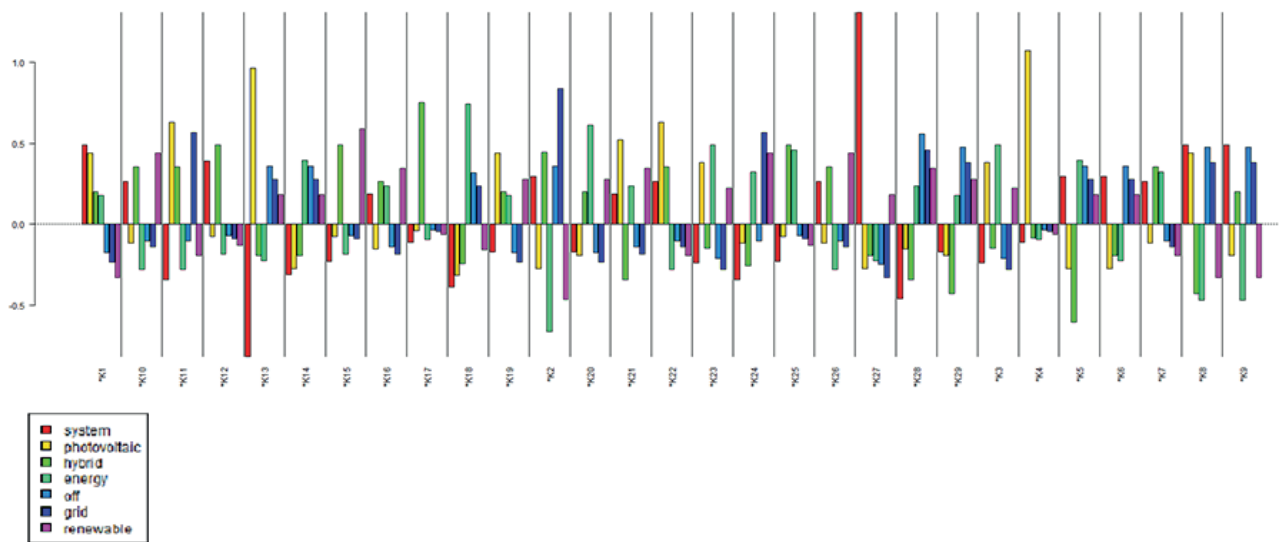


Figura 5. Grafica de Concordancia para las formas con mayor frecuencia. Obtenido del software Iramuteq.

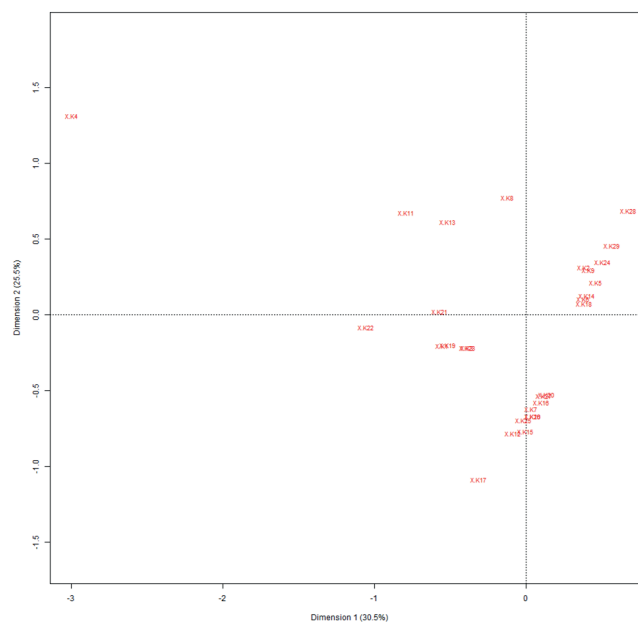


Figura 6. Dendrograma obtenido a partir del Análisis Factorial de Correspondencias, para las variables que representan los grupos de palabras claves. Elaborado con R Commander

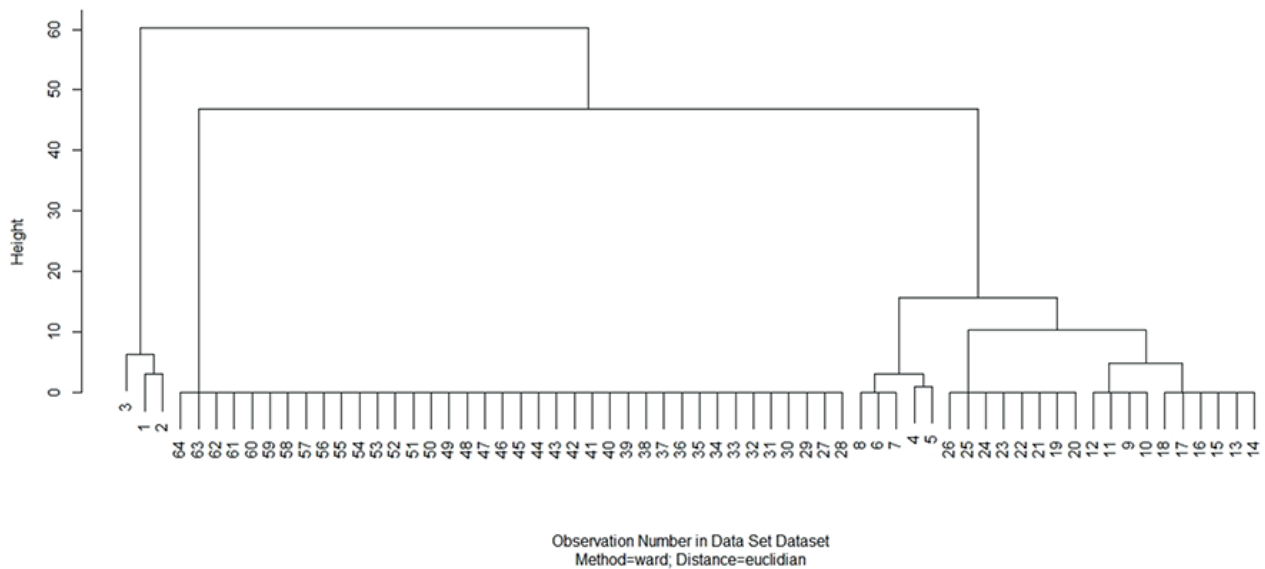


Figura 7. Dendrograma Análisis dimensional de componentes, asociaciones por grupos definidos. Elaborado con R Commander

posibles agrupaciones basadas en las distancias cercanas (agrupamiento jerárquico), el dendrograma identifica las tendencias en agrupaciones de temas, se observa que el grupo 1 y 2 tiene correlación con el grupo 3 y a su vez estos grupos tienen tendencia para asociarse con el bloque de grupos que representan 1 tema en su mayoría, el detalle de estas asociaciones debe ser identificado por el experto en el tema a fin de determinar el conocimiento a adquirir.

5. Discusión

A partir de los resultados se determina que la mayor cantidad de conocimientos generados está alrededor de los off-grid System y hybrid energy systems, la cantidad de temas asociados indica que los nuevos conocimientos y trabajos de investigación tienen su foco de atención en estos temas.

Utilizando Literature Based Discovery [20], [21] se puede determinar a partir de un tema inicial otro distante, a través de un tema intermedio o puente. En nuestro caso a partir de la clusterización basada en asociaciones de palabras clave y text-mining podemos observar que los Off-Grid Hybrid System a pesar de tener cierta antigüedad siguen vigentes en la literatura, lo mismo sucede con Photovoltaic System, respecto a los temas Biomass, Battery capacity degradation, Rural electrification, Photovoltaic with diesel los cuales son tratados en los paper 4, 8, 11 y 13 son considerados temas relevantes, se observa también que el uso de Biomasa para generar energía no es tratado por muchos autores, del mismo modo se muestra que en lugares como Gwakwani y Bangladesh se realizan estudios para el Electricity Access, estas afirmaciones se basan en el análisis de los resultados obtenidos a partir de LBD. Queda al criterio del experto interpretar de forma más precisa las relaciones determinadas a partir de las técnicas de LBD, esto

puede ser discutido pero el sesgo o error en los resultados obtenidos es mínimo, las asociaciones mostradas son válidas y determinan nuevos focos de conocimiento.

6. Conclusiones

Los resultados obtenidos muestran que el análisis de frecuencias es la base para otros análisis, como correspondencias e influencias, de igual forma el análisis de textos se basa en las frecuencias obtenidas de la lematización, lo que permite agrupar palabras o valores correlacionados entre sí a partir de un punto de interés, con esto se generan nuevos grupos lógicos bajo diversos contextos que pueden ser rápidamente inferidos por el experto. Las técnicas de LBD generan correlaciones que no se ven explícitamente, por lo que se necesita algún análisis adicional por parte del experto para determinar correctamente el conocimiento implícito.

References

- [1] Alan L. Porter, Nils C. Newman, Mining external R&D, Technovation, Volume 31, Issue 4, 2011, Pages 171-176, ISSN 0166-4972, <https://doi.org/10.1016/j.technovation.2011.01.001>.
- [2] Kostof R., Briggs M., Solka J. (2007). Literature-related discovery (LRD): Methodology. ScienceDirect-Technological Forecasting & Social Change
- [3] Menasha Thilakarathne, Katrina Falkner, and Thushari Atapattu. 2019. A Systematic Review on Literature-based Discovery: General Overview, Methodology, & Statistical Analysis. ACM Comput. Surv. 52, 6, Article 129 (January 2020), 34 pages. DOI: <https://doi.org/10.1145/3365756>
- [4] Thilakarathne M, Falkner K, Atapattu T. 2019. A systematic review on literature-based discovery workflow. PeerJ Computer Science 5:e235 <https://doi.org/10.7717/peerj-cs.235>
- [5] Henry, S., Panahi, A., Wijesinghe, D. S., & McInnes, B. T. (2019). A Literature Based Discovery Visualization System with Hierarchical Clustering and Linking Set Associations.

- AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science, 2019, 582–591.
- [6] Henry, Samuel & McInnes, Bridget. (2017). Literature Based Discovery: Models, methods, and trends. *Journal of biomedical informatics*. 74. 10.1016/j.jbi.2017.08.011.
- [7] Xu, Z., Yu, D. A Bibliometrics analysis on big data research (2009–2018). *J. of Data, Inf. and Manag.* 1, 3–15 (2019). <https://doi.org/10.1007/s42488-019-00001-2>
- [8] Korhonen A. et al. (2015) Improving Literature-Based Discovery with Advanced Text Mining. In: DI Serio C., Liò P., Nonis A., Tagliaferri R. (eds) *Computational Intelligence Methods for Bioinformatics and Biostatistics. CIBB 2014. Lecture Notes in Computer Science*, vol 8623. Springer, Cham. https://doi.org/10.1007/978-3-319-24462-4_8
- [9] Alberto Moro, Geraldine Joanny, Christian Moretti, Emerging technologies in the renewable energy sector: A comparison of expert review with a text mining software, *Futures*, Volume 117, 2020, 102511, ISSN 0016-3287, <https://doi.org/10.1016/j.futures.2020.102511>
- [10] Trappey, Amy & Chen, Paul & Trappey, Charles & Ma, Lin. (2019). A Machine Learning Approach for Solar Power Technology Review and Patent Evolution Analysis. *Applied Sciences*. 9. 1478. 10.3390/app9071478.
- [11] Porter, a.l., and zhang, y. (2015). "Tech mining of science & Technology information resources for Future-oriented technology analyses," in glenn, j.c. And gordon, t.j. (eds.) *Futures research Methodology version 3.1, the millennium project, washington, dc online at the global futures Intelligence system in futures methodologies*
- [12] Vicente-Gomila, J.M. The contribution of syntactic–semantic approach to the search for complementary literatures for scientific or technical discovery. *Scientometrics* 100, 659–673 (2014). <https://doi.org/10.1007/s11192-014-1299-2>
- [13] Lahlou, Saadi (2001) Text mining methods: an answer to Chartier and Meunier. *Papers on social representations*, 20 (38). pp. 1-7. ISSN 1021-5573
- [14] Sarrica, M., Mingo, I., Mazzara, B., & Leone, G. (2016). The effects of lemmatization on textual analysis conducted with IRaMuTeQ: results in comparison.
- [15] Ramos M.G., do Rosário Lima V.M., Amaral-Rosa M.P. (2019) IRAMUTEQ Software and Discursive Textual Analysis: Interpretive Possibilities. In: Costa A., Reis L., Moreira A. (eds) *Computer Supported Qualitative Research. WCQR 2018. Advances in Intelligent Systems and Computing*, vol 861. Springer, Cham. https://doi.org/10.1007/978-3-030-01406-3_6.
- [16] Sánchez, Evelín & Díaz Blanco, Josval & González, Alicia & Amador Domínguez, Nidia. (2017). Análisis de los Procesos de Lemmatización y Estemizado En Lingüística Computacional. Instituto de Literatura y Lingüística. Cuba. 2017 *Estudios Lingüísticos*, ISBN 978-959-7152-37-8
- [17] Hoffmann, Y., Bisset Alvarez, E., & Martí-Lahera, Y. (2020). Análisis textual con IRaMuTeQ de investigaciones recientes en historia de la educación matemática en Brasil: un ejemplo de Humanidades Digitales. *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, 34(84), 103-133. doi:<http://dx.doi.org/10.22201/ii-bi.24488321xe.2020.84.58097>
- [18] Salone Jean-Jacques. Analyse textuelle avec IRaMuTeQ et interprétations référentielles des programmes officiels de mathématiques en quatrième. Sciences-Croisées, Aix-Marseille Université, 2013. <https://hal.archives-ouvertes.fr/hal-01671161>
- [19] García Moreno, Aitor; Pueyo Mena, Francisco Javier. «Etiquetado gramatical y lematización en el Corpus Histórico Judeoespañol (CORHIJE): problemas, soluciones y resoluciones». *Scriptum digital. Revista de corpus diacrònics i edició digital en Llengües iberoromàniques*, [en línia], 2017, Núm. 6, p. 69-82, <https://www.raco.cat/index.php/scriptumdigital/article/view/329260>.
- [20] Vishrawas Gopalakrishnan, Kishlay Jha, Wei Jin, Aidong Zhang, A survey on literature based discovery approaches in biomedical domain, *Journal of Biomedical Informatics*, Volume 93, 2019, 103141, ISSN 1532-0464, <https://doi.org/10.1016/j.jbi.2019.103141>. (<http://www.sciencedirect.com/science/article/pii/S1532046419300590>).
- [21] Swanson D.R. (2008) Literature-Based Discovery? The Very Idea. In: Bruza P., Weeber M. (eds) *Literature-based Discovery. Information Science and Knowledge Management*, vol 15. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-68690-3_1