

Grafo de conocimiento para determinar el dominio del aprendizaje en la educación superior

Knowledge Graph to determine the domain of learning in Higher Education

Jorge Arturo Hernández-Almazán* | Juan Diego Lumbreras-Vega** | Arturo Amaya Amaya*** | Rubén Machucho-Cadena****

Recepción del artículo: 12/7/2020 | Aceptación para publicación: 25/1/2021 | Publicación: 26/3/2021

RESUMEN

La representación del conocimiento de un estudiante en un área disciplinar juega un rol importante para impulsar sus habilidades. Para apoyar a los involucrados en el ámbito educativo es necesario proporcionarles estrategias de evaluación robustas que faciliten el proceso de enseñanza-aprendizaje. El dominio de un estudiante es determinado por el grado de conocimiento que demuestra, de forma objetiva, sobre los temas incluidos en las diferentes áreas que componen un campo disciplinar. Aunque existe una amplia variedad de técnicas, el grafo de conocimiento en particular está adquiriendo relevancia por el enfoque estructurado y los beneficios que ofrece. Este trabajo propone un método que clasifica y pondera los nodos (temas) de un grafo de conocimiento de un área disciplinar, el cual es analizado mediante un estudio de caso. El método tiene dos enfoques: evitar la evaluación exhaustiva de los nodos y ponderar los nodos con precisión adecuada. Como resultado se obtiene un grafo de conocimiento con sus nodos clasificados y ponderados mediante la aplicación del método propuesto, en el cual 100% de los temas ha sido impactado mediante la evaluación objetiva de 20.8% que representa 10 nodos. Se concluye que el método propuesto tiene potencial para ser utilizado en la representación y la gestión del conocimiento, por lo que es necesario mejorar la iteración de sus fases para condicionar la cantidad de nodos objetivos.

Abstract

The representing of a student's knowledge in an academic discipline plays an important role in boosting the student's skills. To support stakeholders in the educational domain, it is necessary to provide them with robust assessment strategies that facilitate the teaching-learning process. Student's mastery is determined by the degree of knowledge, which demonstrates objectively, on the topics included in the different areas that make up an academic discipline. Although there is a wide variety of techniques to represent knowledge, particularly Knowledge Graph technique is becoming relevant due to the structured approach and benefits it offers. This paper proposes a method that classifies and weights the nodes (topics) of a Knowledge Graph of a disciplinary area, which is analyzed through a case study. The method has two approaches: avoid exhaustive evaluation of the nodes and weight the nodes with adequate precision. Method's application is illustrated by a case study. As results, a Knowledge Graph is obtained with its classified and weighted nodes through the application of the proposed method, in which 100% of the topics have been impacted through the objective evaluation of 20.8% representing 10 nodes. It is concluded that the proposed method has potential to be used in the representation and management of knowledge, being necessary to improve phases' iteration to condition number of objective nodes.



Palabras clave Estructuras de conocimiento; sistemas de clasificación; estudiantes; educación superior



Keywords Knowledge Structures; Classification Systems; Students; Higher Education



INTRODUCCIÓN

El conocimiento se genera a partir de un proceso de transformación de la información (Ramírez & García, 2018), y su representación es esencial en el progreso del ser humano. Existen diferentes técnicas para representar el conocimiento (Han & Ellis, 2020; Lei *et al.*, 2019; Lu *et al.*, 2020; Wang *et al.*, 2019; Yang *et al.*, 2020; Zhang *et al.*, 2020) y una de ellas está atrayendo a los investigadores del ámbito de la educación, la comunicación y de la industria: el grafo de conocimiento (GC) (Li *et al.*, 2019; Mosquera & Piedra, 2018; Yoo & Jeong, 2020; Yu *et al.*, 2017; Zárate *et al.*, 2019; Zhao *et al.*, 2019).

En los últimos años, el GC ha recibido mayor atención en el ámbito educativo (Cope *et al.*,

2020; He *et al.*, 2019; Wang, Ding & Yu, 2019), debido a que se requiere descubrir, medir o compartir el conocimiento de los estudiantes en cada área disciplinar (AD) (Shi *et al.*, 2020). En este sentido, el GC provee una estructura matemática robusta para representar el conocimiento de cualquier AD, por ejemplo, matemáticas, física, química, contaduría, entre otras.

Un AD se compone de gran cantidad de temas que están interrelacionados con base en su complejidad. De acuerdo con el número de temas dominados es posible determinar el grado de conocimiento de un estudiante. Aunque se han hecho esfuerzos por representar el conocimiento de distintas formas (Ivinson, 2020; Paulius & Sun, 2019; Puustinen & Khawaja, 2020), en la actualidad es aún una tarea desafiante que exige procesos



y métodos innovadores (Bloodgood, 2019; Guan *et al.*, 2019; Long *et al.*, 2020) para evitar la evaluación exhaustiva y procurar la precisión.

Por lo tanto, el conocimiento pueden representarse mediante colecciones interrelacionadas de temas en un ámbito determinado. En este sentido, el GC establece relaciones causales entre los temas previos y posteriores de un AD (Rantanen, Hyttinen & Järvisalo, 2020; Shin & Jeong, 2021). Así, la no evaluación exhaustiva se justifica al evitar la evaluación de todos los temas de un AD; es decir, al suponer tres temas interrelacionados por su complejidad $A > B > C$, si al evaluar el tema B se determina 100% de dominio, entonces se asume que el tema A es dominado totalmente, lo que determinaría también parte del conocimiento que se tiene del tema C.

En el presente artículo mostramos los avances iniciales sobre el diseño de un método que clasifica y pondera los nodos (temas) de un GC en un AD. El método tiene dos enfoques: evitar

la evaluación exhaustiva de los nodos y examinar los nodos con precisión adecuada. El método se compone de tres fases que se fundamentan en reglas, las cuales determinan los lineamientos para clasificar y ponderar un nodo de forma objetiva mediante la evidencia proporcionada por el estudiante sobre el dominio de un tema en particular. Después, el grafo es actualizado de forma subjetiva, primero descendente y en seguida ascendente, a partir del efecto causal del nodo antes indicado. Por último, las tres fases son iteradas hasta que el GC es obtenido.

A manera de ilustración, se presenta un estudio de caso que muestra cómo el método propuesto permite obtener un GC del AD de matemáticas. En cada fase del método se considera el enfoque que evita la evaluación exhaustiva de los múltiples temas (nodos) del AD y procura la precisión en la ponderación de los nodos.

El artículo está organizado de la siguiente manera: después de la introducción, se presenta la

revisión de la literatura; posteriormente, se indica la metodología que sustenta el diseño del método propuesto. También se muestran los resultados del estudio de caso y, finalmente, las discusiones y conclusiones.

REVISIÓN DE LA LITERATURA

Hay gran cantidad de investigaciones que aplican el GC para representar el conocimiento (Bellomari *et al.*, 2020; Jia *et al.*, 2017; Krenn & Zeilinger, 2020; Lin *et al.*, 2017; Zhu *et al.*, 2019). La estructura del GC facilita la utilización de ponderaciones que permiten obtener nuevo conocimiento y conclusiones de datos existentes (Chen *et al.*, 2020; Li & Madden, 2019; Wang *et al.*, 2019). Los nodos del GC representan entidades o temas relevantes de un dominio y los bordes (flechas) determinan dependencias, prerrequisitos, relaciones y frecuencia de coocurrencia entre los nodos.

La mayoría de las aplicaciones de GC se enfocan a la extracción de asociaciones de entidades, como *is-a* y *part-of* (Seo *et al.*, 2020; Wang *et al.*, 2019). En contraste, otros estudios consideran como base las diferencias del dominio de los estudiantes para representar, en un GC, las relaciones de prerrequisitos entre los puntos de conocimiento y la recomendación de ejercicios (Lv *et al.*, 2018; Meneses *et al.*, 2020); sin embargo, ellos solo proponen un método que examina los nodos del GC basados en el comportamiento de datos de un estudiante, colectados mediante un sistema de evaluación, pero no utilizan la evidencia proporcionada por el estudiante sobre el dominio de un AD para evitar la exhaustividad en la evaluación.

Oramas *et al.* (2017) y Qiao y Hu (2020) propusieron un método que enriquece la descripción de los nodos en un GC con información semántica. Emplearon dos enfoques diferentes para codificar la información del GC dentro de una representación lineal de las características. Liang *et al.* (2018) destacan el diseño de un esquema de entidad relación para modelar asociaciones

y propiedades de diferentes objetos. A partir del esquema se genera el GC, cuya estructura facilita el aprendizaje en el área del diseño de interiores mediante la ponderación de sus nodos.

Chen *et al.* (2018) proponen un sistema que construye un GC de forma automática en el contexto educativo, combina fuentes de datos heterogéneas para obtener conceptos instruccionales (nodos) e infiere relaciones entre los nodos. Wang *et al.* (2018) representan el conocimiento del AD de geociencias mediante un GC, donde los nodos simbolizan palabras de contenido y la frecuencia de coocurrencia se muestra como borde. En ambos casos, se propone un método compuesto de fases para representar el conocimiento mediante un GC de un AD; cada nodo es etiquetado y ponderado de acuerdo con su influencia entre los temas relevantes. Ninguno de ellos clasifica los nodos para evitar su evaluación exhaustiva como en el caso de Liu *et al.* (2020) y tampoco examina los nodos con una precisión adecuada (Chen *et al.*, 2020).

Las propuestas de Chen *et al.* (2018), Long *et al.* (2020), Shi *et al.* (2020) y Wang *et al.* (2018) se enfocan en integrar fuentes de datos heterogéneas para aplicar principalmente técnicas de procesamiento de lenguaje natural, clasificación o redes neuronales de acuerdo con el contexto del problema; sin embargo, la mayoría evidencia áreas de oportunidad en su aplicación práctica porque la ponderación de los nodos de la red requiere largos tiempos de procesamiento o entrenamiento para representar el conocimiento mediante el GC. Por esta razón, se propone un método que clasifica los nodos del GC que evitan la exhaustividad y los pondera con una precisión adecuada, a partir de la evidencia proporcionada por el estudiante sobre el dominio de los temas de un AD en particular.

METODOLOGÍA

Un AD involucra gran cantidad de temas que son interrelacionados de acuerdo con las dependencias

de sus complejidades, las cuales se pueden representar como nodos y bordes (flechas), respectivamente, en un diagrama llamado grafo del experto. El método propuesto clasifica y pondera los temas (nodos), a partir del grafo del experto, para obtener un GC que representa el dominio de un estudiante sobre los temas de un AD.

En la figura 1 se muestran las tres fases iterativas (proceso) que componen el método propuesto. La entrada para las fases es el grafo del experto, el cual es construido por un experto del AD. En este grafo se establecen las interrelaciones de dependencias mediante bordes (flechas), de acuerdo con la complejidad de los temas (nodos). Así, en la parte superior, es decir, en el primer nivel, se indica el nombre del AD; en el segundo nivel se determinan las áreas del AD; mientras que el tercer nivel presenta los temas (nodo) raíz de las áreas porque son los de menor complejidad.

A partir del tercer nivel hacia abajo se especifican los temas, de esta forma, la complejidad de uno de ellos es mayor entre más alejado se encuentre de su tema (nodo) raíz. Las fases se repiten cada vez que se realiza una evaluación objetiva del estudiante. Después de que las iteraciones de las fases concluyen, se obtiene el GC como salida del método. En la salida de la figura 1 se han resaltado algunos valores y nodos de diferentes tipos,

con el fin de mejorar la apreciación visual. Por lo tanto, el tamaño de estos no es relevante.

Para pasar del grafo del experto (entrada) al GC (salida), los nodos son clasificados en los cinco tipos mostrados en la tabla 1. Por lo tanto, los nodos que aparecen en el GC de la figura 1 son definidos en la tabla 1.

Cada tipo de nodo, excepto el V, tendrá una ponderación actual (PA) que representa el dominio de un estudiante sobre el tema en el GC. Los nodos O y S pueden tener una PA de 0 a 100. La PA del nodo O se determina a partir de la evaluación objetiva de un estudiante mediante un cuestionario con n preguntas o problemas sobre el tema del nodo O. Por lo tanto, la PA del nodo O es el porcentaje de aciertos que un estudiante ha obtenido de la evaluación objetiva. Por esta razón, cada estudiante tendrá un GC diferente en función de su dominio del AD.

Los nodos de tipo SV y SL tendrán una PA constante de 3 y 2, respectivamente, mientras que el nodo V no tiene PA porque forma parte del grafo del experto. Así, el cálculo de la PA se realiza en las tres fases del método, es decir, en el proceso mostrado en la figura 1.

En seguida se detallan las tres fases del método propuesto y las reglas que las fundamentan.

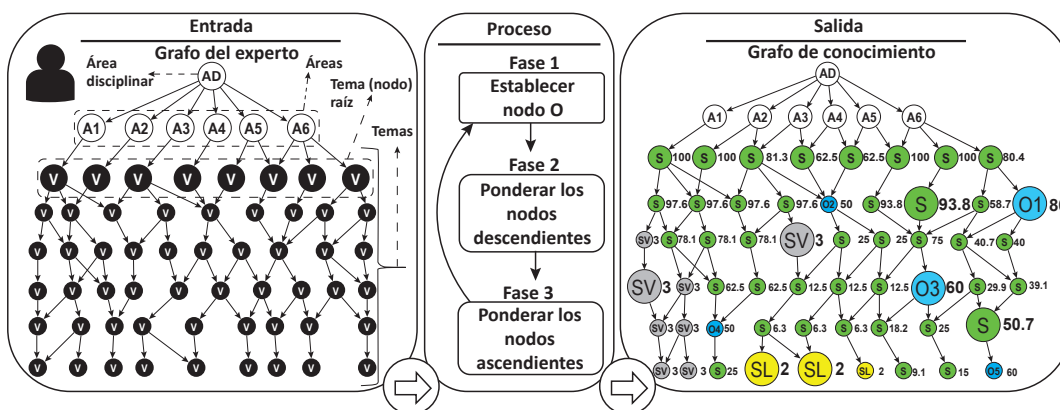


Figura 1. Esquema general del método propuesto.
Fuente: elaboración propia.

Tabla 1. Clasificación de los nodos del GC

Tipo	DESCRIPCIÓN
V	Nodo virgen: todos los nodos del grafo del experto son del tipo V y no tienen ponderación. Este tipo de nodo tiene la menor prioridad en el grafo. Es de color negro
SV	Nodo subjetivo virgen: en la fase 1 los nodos se convierten al tipo SV, cada uno con ponderación de 3 para representar un grado de dominio bajo. El nodo SV sustituye a un nodo de tipo V. Es de color gris
O	Nodo objetivo: es determinado cuando se obtiene la evidencia del dominio del tema (nodo) por parte de un estudiante mediante un cuestionario con un valor mínimo de 0 y máximo de 100. El nodo O impactará a sus nodos interrelacionados de forma ascendiente y descendiente. Este tipo de nodo tiene la mayor prioridad en el grafo; por lo tanto, puede sustituir a un nodo de cualquier tipo: SV, S y SL. Es de color azul
S	Nodo subjetivo: se genera a partir del efecto causal del nodo O. Un nodo S forma parte de los nodos ascendientes y descendientes que están interrelacionados al nodo O. El valor del nodo S es calculado subjetivamente a partir del valor del nodo O. El nodo S puede sustituir a un nodo de tipo SV o SL. Es de color verde
SL	Nodo subjetivo lejano: se genera a partir del efecto causal del nodo O. Un nodo SL forma parte de los nodos descendientes que están interrelacionados al nodo O a partir del cuarto nivel de complejidad hacia abajo. Debido a que un nodo SL se encuentra lejano del nodo O, su ponderación tiene más incertidumbre que el nodo S; por lo tanto, el nodo SL tendrá una ponderación constante de 2 para representar un grado de dominio bajo. El nodo SL puede sustituir a un nodo de tipo SV. Es de color amarillo

Fuente: elaboración propia.

Fase 1. Establecer nodo O

En esta fase se determina un nodo del tipo O al que se le asocia la PA obtenida de la evaluación objetiva del estudiante. Las reglas de esta fase son:

F1.R1) El nodo O sustituye a un nodo de tipo SV, S y SL.

F1.R2) Un nodo O se establece aleatoriamente sobre un nodo que se encuentre debajo de los primeros dos niveles del grafo (el primer nivel es el AD y el segundo nivel son las áreas del AD). Además, los tres primeros nodos O deberán estar antes de los dos últimos niveles del grafo (los temas más complejos). Esta regla asegura que los nodos O se establezcan exclusivamente en un tema (no en el nodo que representa el AD ni en los nodos que representan las áreas del AD). Además, asegura que los tres primeros nodos O presenten temas de mediana complejidad.

F1.R3) Un nodo O no puede ser sustituido por ningún tipo de nodo.

F1.R4) Un nodo O es evaluado objetivamente una sola ocasión.

A continuación, se describen los pasos de la fase 1, los cuales respetan las reglas mencionadas:

F1.P1) Los nodos V se convierten al tipo SV, cada uno con ponderación de 3 para representar un grado de dominio bajo.

F1.P2) Se elige un nodo aleatoriamente.

F1.P3) Si el nodo elegido es de tipo S, SL o SV, se sustituye por el nodo O. De lo contrario, regresa al paso F1.P2.

F1.P4) Al nodo O se le asigna la PA obtenida de la evaluación objetiva del estudiante. De esta forma, se obtiene la evidencia proporcionada por el estudiante sobre el dominio de un tema (nodo) en particular.

F1.P5) El color del nodo O se establece azul.

En la figura 2 se muestra un ejemplo de referencia de cómo se actualiza el grafo en cada fase durante una iteración. En la figura 2 (a) se muestra el estado inicial del grafo del experto (entrada). En la figura 2 (b) se aprecia la transición cuando el nodo O ha sido establecido con una ponderación de 80 (PA=80), es decir, cuando la

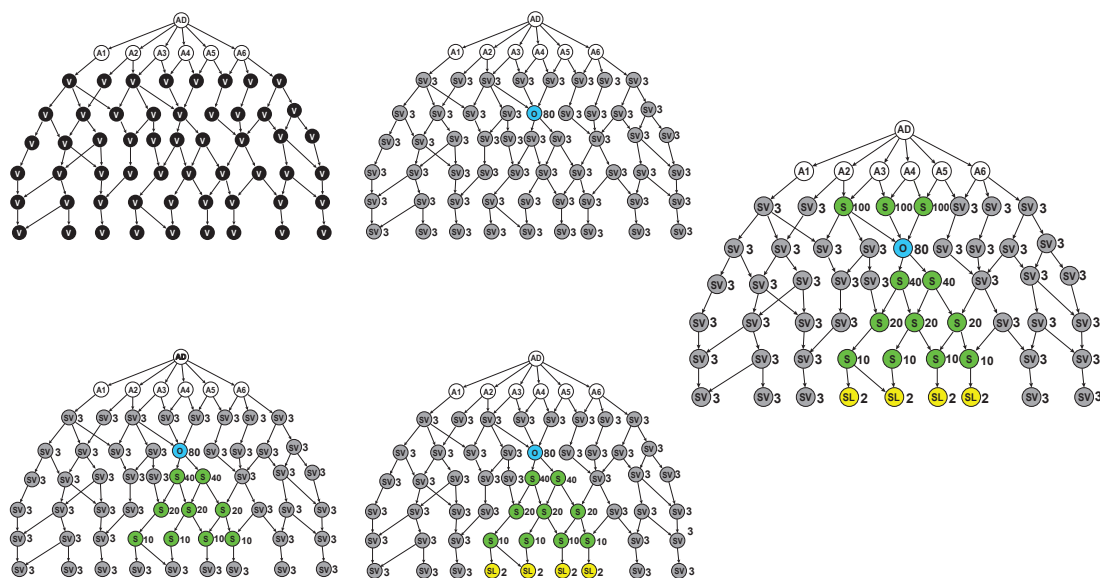


Figura 2. Actualización del grafo a través de las tres fases del método: ejemplo de una iteración.

Fuente: elaboración propia.

fase 1 ha concluido. Los valores mostrados son la PA de cada nodo.

En las reglas de la fase 2 y fase 3 se especifican valores y porcentajes que hemos propuesto para que el método evite la evaluación exhaustiva de los nodos y realice la ponderación con una precisión adecuada. Aunque esta es una propuesta de los autores, se ha tomado en cuenta la relación causal entre los diferentes niveles de conocimiento en un AD con su grado de complejidad respecto a los conocimientos previos y posteriores (Nie, Shi & Li, 2020; Rantanen, Hyttinen & Järvisalo, 2020; Shin & Jeong, 2021).

Fase 2. Ponderar los nodos descendientes

En esta fase se ponderan los nodos descendientes que están interrelacionados al nodo O. Así, el efecto causal del nodo O impacta a los nodos (temas) con mayor complejidad que el mismo. Los nodos impactados serán sustituidos por nodos del tipo S o SL. Las reglas de la fase 2 son:

F2.R1) El efecto causal del nodo O impacta a sus nodos hijos, nietos y bisnietos. En otras palabras, son los nodos que están a máximo tres saltos del nodo O. Estos nodos son clasificados como tipo S.

F2.R2) Los nodos S clasificados en el paso anterior, reciben una ponderación de 50% de su padre. Por lo tanto, un hijo del nodo O recibe 50% de la PA del nodo O. Un nieto del nodo O recibe 50% de la PA de su padre (hijo del nodo O). Un bisnieto del nodo O recibe 50% de la PA de su padre (nieto del nodo O). Esta regla se debe a que los nodos S, antes indicados, representan temas más complejos que el tema del nodo O, por consecuencia a partir de la evidencia del nodo O el método asume que el estudiante domina el nodo siguiente en complejidad (hijo del nodo O) en 50% de la PA del nodo O, y así sucesivamente.

F2.R3) La PA de un nodo S es el promedio de las ponderaciones que recibe sucesivamente. En la ecuación 1 se muestra la expresión para calcular la PA de un nodo S:

$$PA = \frac{\sum P}{NP} \quad (1)$$

Donde P son las ponderaciones recibidas en el nodo S y NP es el número de ponderaciones recibidas en el nodo S.

F2.R4) El efecto causal del nodo O impacta a los nodos posteriores de sus bisnietos. En otras palabras, son los nodos que están a partir del cuarto salto del nodo O. Estos nodos son clasificados como tipo SL.

F2.R5) La PA de un nodo SL es de 2 porque se asume que el estudiante tiene un grado de dominio bajo de este tema (nodo). Aunque el nodo SL está interrelacionado con el nodo O, el efecto causal del nodo O no impacta al nodo SL tanto como al nodo S (con 50% de la PA de su padre). Lo anterior se justifica por la distancia lejana del nodo SL con respecto al nodo O.

A continuación, se describen los pasos de la fase 2, los cuales respetan las reglas mencionadas:

F2.P1) Se determinan los caminos que tienen nodos interrelacionados, a máximo tres saltos, al nodo O descendientemente. Un camino puede tener nodos que han sido considerados en otros caminos.

F2.P2) Los caminos son ordenados de manera ascendente de acuerdo con su número de nodos. De esta forma, el camino con menos nodos es el primero en el orden.

F2.P3) Si existen n caminos que tienen el mismo número de nodos, los n caminos son ordenados de manera descendente de acuerdo con el número de nodos S que tiene cada uno de los n caminos. De esta forma, el camino con más nodos S será el primero en el orden de los n caminos.

F2.P4) Si existen n caminos que tengan el mismo número de nodos y el mismo número de nodos S no se realiza ordenamiento en los n caminos.

F2.P5) Se obtiene la PA de los nodos interrelacionados en cada uno de los caminos que han sido ordenados mediante los cuatro pasos anteriores. El primer camino para ponderar los nodos es el primero en el orden obtenido, y así sucesivamente. En cada camino se pondera primero el hijo, luego el nieto y por último el bisnieto, todos con respecto al nodo O, según corresponda. Esta ponderación se fundamenta en las reglas F2.R1 a la F2.R3.

F2.P6) El color del nodo S se establece en verde.

F2.P7) Se determinan los nodos SL interrelacionados al nodo O descendientemente. Los nodos SL están a partir del cuarto salto del nodo O.

F2.P8) A cada nodo SL se le asigna una PA de 2.

F2.P9) El color del nodo SL se establece amarillo.

En la figura 2 (c) y (d) se muestran los nodos S y SL, respectivamente, clasificados y ponderados al finalizar la fase 2. Los valores mostrados son la PA de cada nodo.

Si existen n caminos que tienen el mismo número de nodos, los n caminos son ordenados de manera descendente de acuerdo con el número de nodos S que tiene cada uno de los n caminos. De esta forma, el camino con más nodos S será el primero en el orden de los n caminos

El primer camino para ponderar los nodos es el primero en el orden obtenido, y así sucesivamente. En cada camino se pondera primero el padre y luego el abuelo, ambos con respecto al nodo O, y así sucesivamente hasta llegar al tema (nodo) raíz

Fase 3. Ponderar los nodos ascendientes

En esta fase se ponderan los nodos ascendientes que están interrelacionados al nodo O. Así, el efecto causal del nodo O impacta a los nodos (temas) con menor complejidad que el mismo. Los nodos impactados serán sustituidos por nodos del tipo S. Las reglas de la fase 3 son:

F3.R1) Un tema (nodo) raíz se encuentra después del segundo nivel del grafo; por lo tanto, es hijo de un AD.

F3.R2) El efecto causal del nodo O impacta a sus nodos padres, abuelos y así sucesivamente hasta llegar al tema (nodo) raíz. En otras palabras, son los temas (nodos) interrelacionados que anteceden al nodo O. Estos nodos son clasificados como tipo S.

F3.R3) Los nodos S clasificados en el paso anterior reciben una ponderación de 25% más de la PA de su hijo. En la ecuación 2 se muestra la expresión para calcular la ponderación que recibe un nodo S en la fase 3.

$$PR = PA_S * 1.25 \quad (2)$$

Donde PR es una ponderación recibida en el nodo S durante la fase 3 y PA es la ponderación actual del hijo del nodo S. Esta regla se debe a que los nodos S en la fase 3 representan temas menos complejos que el tema del nodo O; por consecuencia, a partir de la evidencia del nodo O se asume que el estudiante domina el nodo anterior en complejidad (padre del nodo O) en 25% más que la PA del nodo O, y así de manera sucesiva.

F3.R4) La PA de un nodo S es el promedio de las ponderaciones que recibe de manera consecutiva. En la ecuación 1 se muestra la expresión para calcular la PA de un nodo S.

F3.R5) Si la PA del nodo S es mayor que 100, la PA se actualiza a 100 para mantener el límite máximo de ponderación.

A continuación, se describen los pasos de la fase 3, los cuales respetan las reglas mencionadas:

F3.P1) Se determinan los caminos que tienen nodos ascendientes a partir del nodo O hasta el tema (nodo) raíz. Un camino puede tener nodos que han sido considerados en otros caminos.

F3.P2) Se realizan los pasos F2.P2, F2.P3 y por último el F2.P4.

F3.P3) Se obtiene la PA de los nodos interrelacionados en cada uno de los caminos que han sido ordenados mediante los pasos F3.P1 y F3.P2. El primer camino para ponderar los nodos es el primero en el orden obtenido, y así sucesivamente. En cada camino se pondera primero el padre y luego el abuelo, ambos con respecto al nodo O, y así de manera consecutiva hasta llegar al tema (nodo) raíz, según corresponda. Esta ponderación se fundamenta de las reglas F3.R1 a la F3.R5.

F3.P4) Se realiza el paso F2.P6.

En la figura 2 (e) se muestran los nodos S clasificados y ponderados al finalizar la fase 3. Los valores mostrados son la PA de cada nodo.

Al terminar la fase 3, se ha concluido una iteración. En seguida, se itera a la fase 1 para establecer el siguiente nodo O. Después de concluir las iteraciones de las fases, se obtiene el GC que representa el dominio del estudiante sobre los temas de un AD.

ESTUDIO DE CASO

Como ya se mencionó, el dominio educativo necesita representar el conocimiento de los estudiantes mediante procesos y métodos innovadores (Guan *et al.*, 2019; Long *et al.*, 2020) que eviten la evaluación exhaustiva (Liu *et al.*, 2020) y procuren la precisión (Chen *et al.*, 2020). El presente estudio de caso describe cómo se puede estudiar este problema de manera simplificada.

Primero, solicitamos a un experto en matemáticas que definiera un grafo de esa AD. El grafo del experto se compone de las siguientes seis áreas: aritmética, álgebra, probabilidad y estadística, geometría, trigonometría y cálculo, las cuales incluyen un total de 48 temas (nodos) interrela-

cionados de acuerdo con su complejidad. Así, el grafo del experto integra solo algunos temas de matemáticas. Los temas representan los nodos del tipo V. En segundo lugar, usamos este grafo del experto como línea base para aplicar diez iteraciones de las tres fases del método propuesto. El estado del grafo, al terminar la iteración 1, es la entrada para la fase 1 de la iteración 2, y así de manera sucesiva.

Finalmente, se establecen diez nodos O, uno por cada iteración, para obtener el GC. Las PA de los diez nodos O se han dispuesto de forma arbitraria al suponer a un estudiante simulado.

RESULTADOS

En la tabla 2 se especifican los resultados de las diez iteraciones, el nodo O establecido en la fase 1, la evidencia proporcionada por el estudiante sobre el dominio del tema (nodo O) del AD de matemáticas y los porcentajes de cada tipo de nodo al terminar cada iteración. Al concluir la iteración 10 se obtiene el GC mostrado en la figura 3.

En la tabla 2 se observa que, al finalizar la iteración 1, solo se registró 2.1% de nodos SL, lo cual representa a un nodo; esto se debe a que la estadística

Tabla 2. Resultados de las iteraciones del método

ITERACIÓN	Nodo O	% DE DOMINIO DEL ESTUDIANTE	% DE NODOS EN EL GRAFO			
			O	S	SL	SV
1	Variables	34	2.1	35.4	2.1	60.4
2	Radicales	17	4.2	52.1	8.3	35.4
3	Triángulos	63	6.3	62.5	20.8	10.4
4	Multiplicación y división	80	8.3	66.7	20.8	4.2
5	Seno y coseno	87	10.4	77.1	12.5	0.0
6	Hipérbola	72	12.5	77.1	10.4	0.0
7	Derivadas	91	14.6	75.0	10.4	0.0
8	Distribución binomial	76	16.7	72.9	10.4	0.0
9	Factorización	58	18.8	81.3	0.0	0.0
10	Recta	43	20.8	79.2	0.0	0.0

Fuente: elaboración propia.

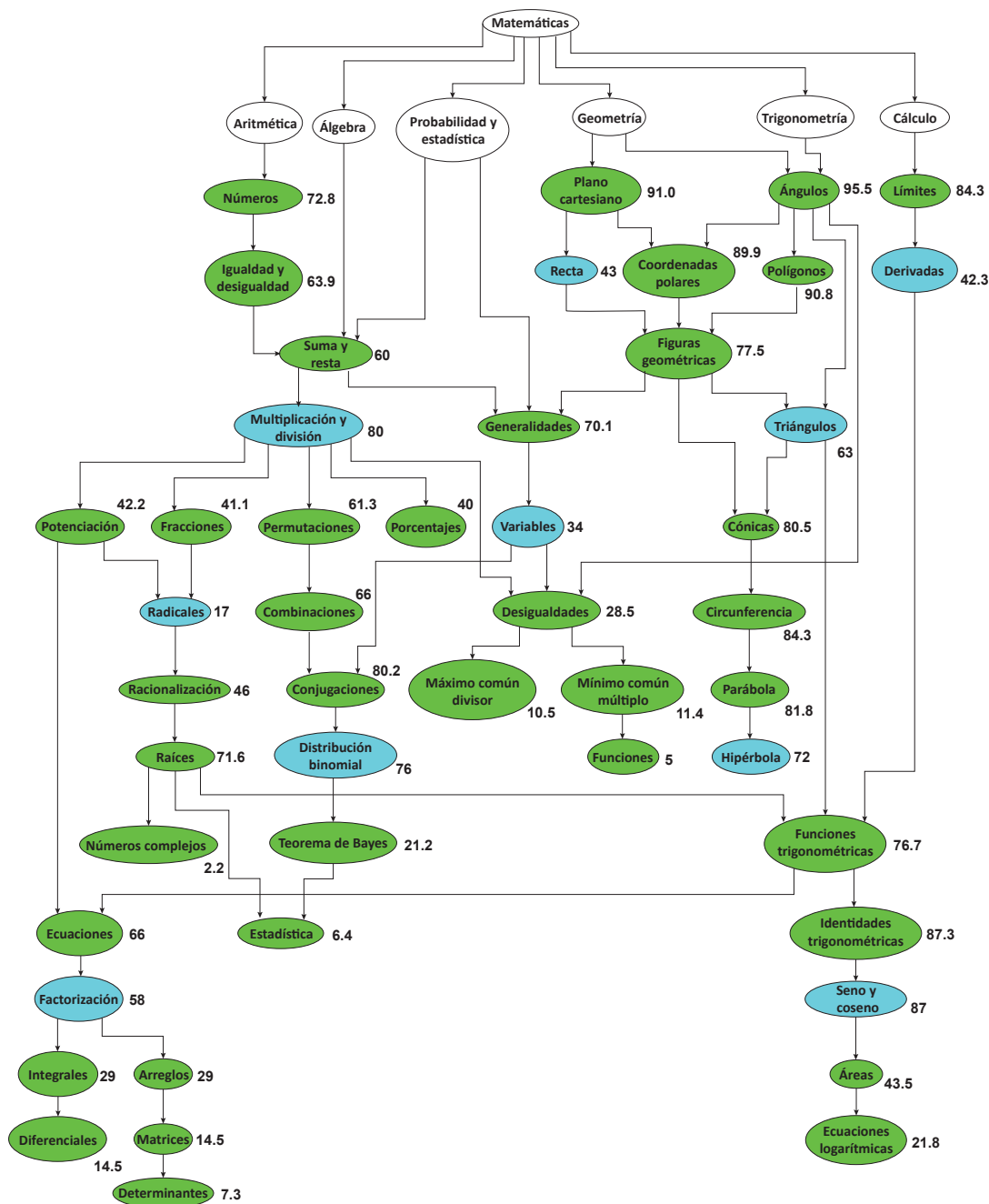


Figura 3. Segmento del grafo de conocimiento.
Fuente: elaboración propia.

es el único nodo a partir del cuarto salto del nodo O variables (figura 3). Además, la PA del nodo O Variables permite clasificar y ponderar 18 nodos interrelacionados de manera subjetiva, a partir de la primera evaluación objetiva sobre el dominio del tema del estudiante. Este enfoque evita la evaluación exhaustiva de los temas del GC y permite ponderar los nodos de forma adecuada porque se fundamenta en la evaluación objetiva, es decir, la PA del nodo O.

Al concluir, se ha establecido un nodo O por cada iteración. Como resultado, la evaluación de los primeros cinco nodos O ha permitido clasificar y ponderar 100% de los nodos del GC, de los cuales 89.6% ha sido impactado por el efecto causal de los nodos O. En este sentido, al finalizar la iteración del quinto nodo O (Seno y coseceno) no existen nodos SV en el GC. La iteración 9 de la tabla 2 muestra la eliminación de los nodos SL, mientras que en la última iteración se especifica que la evaluación objetiva de 20.8% de los nodos del grafo ha impactado de forma subjetiva al resto de los nodos, como se ilustra en la figura 3. Por lo tanto, la figura 3 es el GC que representa el dominio de un estudiante sobre los temas del AD de matemáticas.

Aunque el método propuesto permite clasificar y ponderar los nodos del GC de forma sistemática y acorde a la complejidad de los temas, pareciera que la figura 3 no demuestra la precisión adecuada. Por ejemplo, el nodo Conjunciones tiene una PA mayor que la de sus dos nodos padres combinaciones y variables. En otras palabras, podría resultar confuso porque el estudiante tiene más dominio en un tema (nodo) de mayor complejidad que en un tema (nodo) de menor complejidad.

Lo anterior se debe a que el nodo Conjunciones fue impactado por el efecto causal de tres nodos O. El primero, nodo

Al concluir, se ha establecido un nodo O por cada iteración. Como resultado, la evaluación de los primeros cinco nodos O ha permitido clasificar y ponderar 100% de los nodos del GC

Variables, tiene una PA baja (34), mientras que el segundo, nodo Multiplicación y división, y el tercero, nodo Distribución binomial, tienen una PA mayor, 80 y 76, respectivamente. Por lo tanto, la PA del nodo Conjunciones (nodo S) es actualizada una vez por cada uno de los caminos que lo contemplan, los cuales se determinan en la fase 2 y fase 3 del método propuesto.

El método se enfoca en clasificar solamente temas. Por esta razón, en la figura 3 el nodo del primer nivel (AD) y los seis nodos del segundo nivel (áreas de la AD), no son clasificados ni ponderados en el GC obtenido. Por último, la cantidad de los tipos de nodos S y SL en cada iteración depende del nodo O establecido. Por lo tanto, si se realizan más evaluaciones con estudiantes diferentes, cada estudiante tendrá una instancia independiente del GC de acuerdo con su grado de dominio de los temas de la AD.

DISCUSIÓN Y CONCLUSIONES

Aunque los GC son ampliamente utilizados y aplicados en investigaciones actuales de diversas áreas, todavía queda trabajo por hacer en el ámbito educativo. En este artículo presentamos nuestra contribución

inicial mediante un método que clasifica y pondera los nodos (temas) de un grafo de conocimiento de un área disciplinar. En el método se toma como línea base el grafo del experto para que sea la entrada de las tres fases iterativas. Al finalizar las iteraciones, se obtiene el GC que representa el dominio de un estudiante sobre los temas del AD.

La evaluación exhaustiva de los nodos del GC se evita mediante el efecto causal del nodo tipo O, el cual impacta a los nodos ascendientes y descendientes que tiene interrelacionados. Además de que los nodos interrelacionados son clasificados al tomar ventaja de la evaluación objetiva del nodo tipo O, también son ponderados con una precisión adecuada al considerar como base la PA del nodo tipo O. Para ser justos, el estudio de caso describe la prueba del método con la evidencia de un estudiante simulado. Nuestro propósito es mostrar la aplicación de cada fase del método con sus respectivas reglas para exponer su potencial en el dominio educativo.

El estudio de caso ha demostrado que el método propuesto evita la evaluación exhaustiva de los nodos (temas) de la AD, y consigue una ponderación adecuada. En este sentido, al evaluar objetivamente solo 20.8% de los nodos, el método ha permiti-

tido obtener un GC que manifiesta el grado de dominio del estudiante sobre el AD. Con base en este antecedente, el método propuesto puede ser utilizado en otros escenarios educativos. Por ejemplo, en el nivel medio superior y superior en AD como química, física, contabilidad, entre otras. Para hacerlo, se recomienda seguir las fases y reglas establecidas previamente. Sin embargo, en AD complejas y con temas intrínsecamente interrelacionados, se elevaría la dificultad de realizar la ponderación y clasificación de los nodos; esto se debe porque al seguir las reglas mencionadas se tendría que determinar los n caminos que tienen los nodos interrelacionados al nodo O.

Hay varios aspectos por mejorar en este método y su implementación. Primero, el proceso de iteración de las fases aún no es sólido porque no se especifica una pauta de cuántos nodos O establecer. Como un método que evita la evaluación exhaustiva, esperamos implementar un módulo inteligente basado en redes bayesianas que infiere cuándo finalizar las iteraciones. Integraremos diferentes fuentes de datos heterogéneas que mejoren la precisión de la ponderación de los nodos. Por ejemplo, los resultados del examen de los estudiantes antes de ingresar a la universidad y su comportamiento en las plataformas tecnológicas que gestionan el aprendizaje.

En segundo lugar, debido a que cada estudiante tendrá un GC diferente en función de su dominio del AD, es necesario automatizar el proceso mediante un sistema que gestione el conocimiento de los estudiantes sobre un AD, el cual se base en el enfoque del método propuesto. En esta misma dirección, planeamos desarrollar módulos dirigidos a cuatro tipos de usuarios finales: experto, estudiante, profesor y administrador.

La evaluación exhaustiva de los nodos del GC se evita mediante el efecto causal del nodo tipo O, el cual impacta a los nodos ascendientes y descendientes que tiene interrelacionados

En tercer lugar, el método propuesto puede potenciar el aprendizaje en la educación universitaria debido a que el GC representa el dominio de un estudiante sobre un AD. Una vez que se conoce el grado de conocimiento del estudiante, se puede favorecer al desarrollo de sus competencias disciplinarias y profesionales mediante la atención puntual de los temas (nodos) con deficiencia.

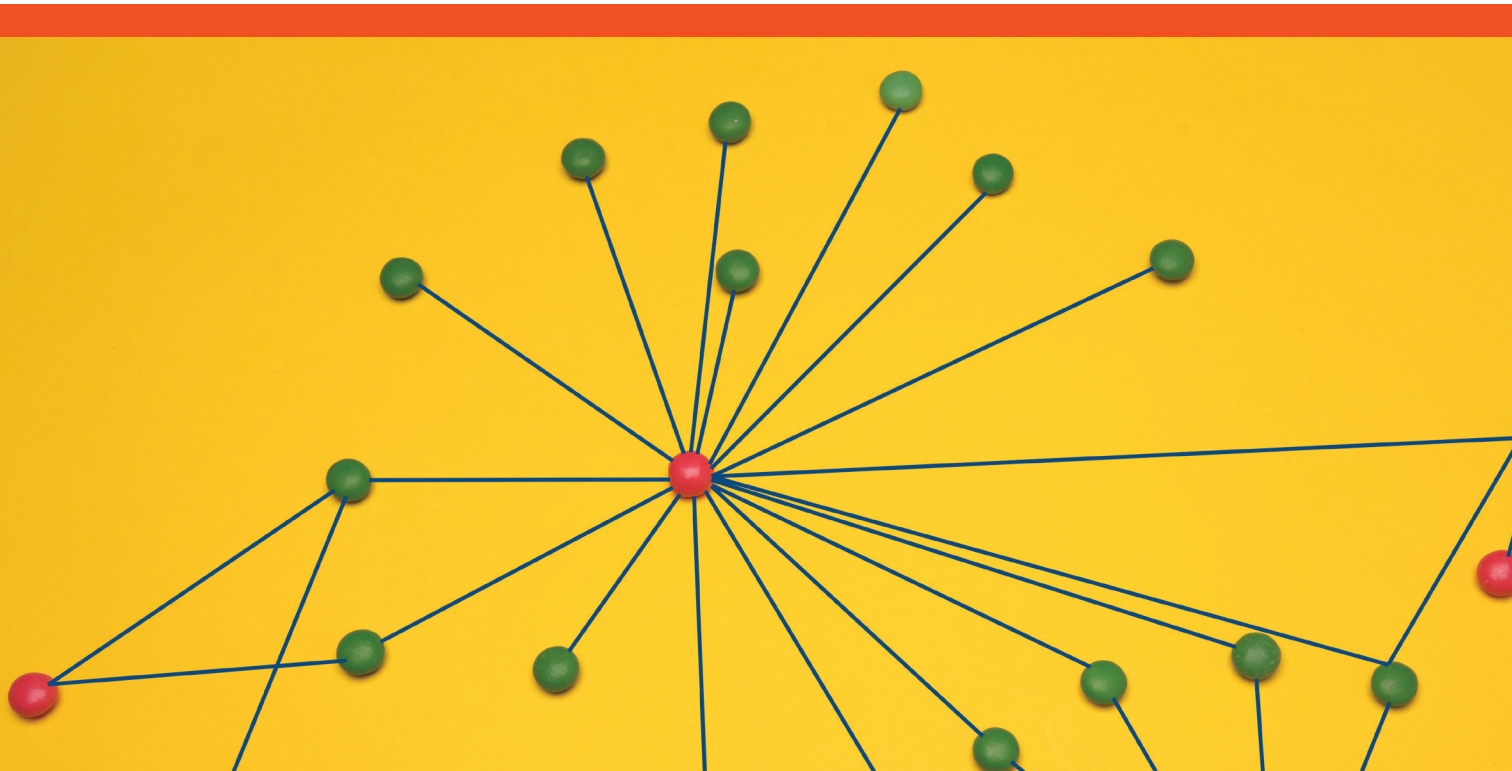
Finalmente, pretendemos investigar las formas más efectivas de determinar rutas de aprendizaje de acuerdo con el GC de cada estudiante a partir de las recomendaciones. En particular, la gestión de conocimiento mediante un GC orientado al dominio educativo es nuestra perspectiva de investigación. *a*

REFERENCIAS

- Bellomarini, L.; Benedetto, D.; Gottlob, G. & Sallinger, E. (2020). Vadalog: A modern architecture for automated reasoning with large knowledge graphs. *Information Systems*. <https://doi.org/10.1016/j.is.2020.101528>
- Bloodgood, J. M. (2019). Knowledge acquisition and firm competitiveness: the role of complements and knowledge source. *Journal of Knowledge Management*, 23(1), 46-66. <https://doi.org/10.1108/jkm-09-2017-0430>
- Chen, P.; Lu, Y.; Zheng, V. W.; Chen, X. & Yang, B. (2018). Knowedu: A system to construct knowledge graph for education. *IEEE Access*, 6, 31553-31563. <https://doi.org/10.1109/access.2018.2839607>
- Chen, W.; Jiang, M.; Jiang, C. & Zhang, J. (2020). Critical node detection problem for complex network in undirected weighted networks. *Physica A: Statistical Mechanics and its Applications*, 538. <https://doi.org/10.1016/j.physa.2019.122862>
- Chen, X.; Jia, S. & Xiang, Y. (2020). A review: Knowledge reasoning over knowledge graph. *Expert Systems with Applications*, 141. <https://doi.org/10.1016/j.eswa.2019.112948>
- Cope, B.; Kalantzis, M. & Sears, D. (2020). Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. *Educational Philosophy and Theory*, 1-17. <https://doi.org/10.1080/00131857.2020.1728732>
- Guan, N.; Song, D. & Liao, L. (2019). Knowledge graph embedding with concepts. *Knowledge-Based Systems*, 164, 38-44. <https://doi.org/10.1016/j.knsys.2018.10.008>
- Han, F. & Ellis, R. (2020). Redes de aprendizaje personalizadas en contextos universitarios de aprendizaje semipresencial. *Comunicar*, 28(62), 19-30. <https://doi.org/10.3916/c62-2020-02>
- He, M.; Du, X. & Wang, B. (2019). Representation learning of Knowledge Graphs via fine-grained relation description combinations. *IEEE Access*, 7, 26466-26473. <https://doi.org/10.1109/access.2019.2901544>
- Iverson, G. (2020). The power of living knowledge: re-imagining horizontal knowledge. *Asia-Pacific Journal of Teacher Education*, 48(1), 15-29. <https://doi.org/10.1080/1359866x.2019.1696453>
- Jia, Y.; Wang, Y.; Jin, X.; Lin, H. & Cheng, X. (2017). Knowledge graph embedding: A locally and temporally adaptive translation-based approach. *ACM Transactions on the Web*, 12(2), 1-33. <https://doi.org/10.1145/3132733>
- Krenn, M. & Zeilinger, A. (2020). Predicting research trends with semantic and neural networks with an application in quantum physics. *Proceedings of the National Academy of Sciences of the United States of America*, 117(4), 1910-1916. <https://doi.org/10.1073/pnas.1914370116>
- Lei, Y.; Yucong, D. & Yonghao, Z. (2019). Knowledge graph provision for heterogeneous service network. 2019 IEEE World Congress on Services (SERVICES), Milan, Italy.
- Li, D. & Madden, A. (2019). Cascade embedding model for knowledge graph inference and retrieval. *Information Processing and Management*, 56(6). <https://doi.org/10.1016/j.ipm.2019.102093>
- Li, Y.; Cao, J. & Wang, Y. (2019). Implementation of intelligent question answering system based on basketball knowledge graph. 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China.

- Liang, Y.; Xu, F.; Zhang, S.-H.; Lai, Y.-K. & Mu, T. (2018). Knowledge graph construction with structure and parameter learning for indoor scene design. *Computational Visual Media*, 4(2), 123-137. <https://doi.org/10.1007/s41095-018-0110-3>
- Lin, Z.-Q.; Xie, B.; Zou, Y.-Z.; Zhao, J.-F.; Li, X.-D.; Wei, J.; Sun, H.-L. & Yin, G. (2017). Intelligent development environment and software knowledge graph. *Journal of Computer Science and Technology*, 32(2), 242-249. <https://doi.org/10.1007/s11390-017-1718-y>
- Liu, S.; Ding, C.; Ma, Y.; Li, Z. & Xiang, J. (2020). Redundancy reduction based node classification with attribute augmentation. *Knowledge-Based Systems*, 188. <https://doi.org/10.1016/j.knsys.2019.105080>
- Long, J.; Chen, Z.; He, W.; Wu, T. & Ren, J. (2020). An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market. *Applied Soft Computing*, 91. <https://doi.org/10.1016/j.asoc.2020.106205>
- Lu, R.; Fei, C.; Wang, C.; Gao, S.; Qiu, H.; Zhang, S. & Cao, C. (2020). HAPE: A programmable big knowledge graph platform. *Information Sciences*, 509, 87-103. <https://doi.org/10.1016/j.ins.2019.08.051>
- Lv, P.; Wang, X.; Xu, J. & Wang, J. (2018). Utilizing knowledge graph and student testing behavior data for personalized exercise recommendation. TURC 2018: ACM Turing Celebration Conference, China.
- Meneses Ortégón, J. P.; Jové, T.; Puiggalí, J. & Fabregat, R. (2020). Representación del conocimiento de un proceso de co-creación de material educativo. *Tecnológicas*, 23(47), 159-177. <https://doi.org/10.22430/22565337.1493>
- Mosquera, J. & Piedra, N. (2018). Use of graph database for the integration of heterogeneous data about ecuadorian historical personages. 2018 7th International Conference On Software Process Improvement (CIMPS), Guadalajara, Mexico.
- Nie, F.; Shi, S. & Li, X. (2020). Semi-Supervised Learning with Auto-Weighting Feature and Adaptive Graph. *IEEE Transactions on Knowledge and Data Engineering*, 32(6), 1167-1178. <https://doi.org/10.1109/tkde.2019.2901853>
- Oramas, S.; Ostuni, V. C.; Noia, T. D.; Serra, X. & Sciascio, E. D. (2017). Sound and music recommendation with knowledge graphs. *ACM Transactions on Intelligent Systems and Technology*, 8(2), 1-21. <https://doi.org/10.1145/2926718>
- Paulius, D. & Sun, Y. (2019). A survey of knowledge representation in service robotics. *Robotics and Autonomous Systems*, 118, 13-30. <https://doi.org/10.1016/j.robot.2019.03.005>
- Puustinen, M. & Khawaja, A. (2020). Envisaging the alternatives: from knowledge of the powerful to powerful knowledge in history classrooms. *Journal of Curriculum Studies*, 1-16. <https://doi.org/10.1080/00220272.2019.1703273>
- Qiao, C. & Hu, X. (2020). A neural knowledge graph evaluator: Combining structural and semantic evidence of knowledge graphs for predicting supportive knowledge in scientific QA. *Information Processing & Management*, 57(6). <https://doi.org/10.1016/j.ipm.2020.102309>
- Rantanen, K.; Hyttinen, A. & Järvisalo, M. (2020). Discovering causal graphs with cycles and latent confounders: An exact branch-and-bound approach. *International Journal of Approximate Reasoning*, 117, 29-49. <https://doi.org/10.1016/j.ijar.2019.10.009>
- Ramírez-Montoya, M.-S. & García-Peñalvo, F.-J. (2018). Co-creación e innovación abierta: Revisión sistemática de literatura. *Comunicar*, 26(54), 09-18. <https://doi.org/10.3916/c54-2018-01>
- Seo, S.; Oh, B. & Lee, K.-H. (2020). Reliable knowledge graph path representation learning. *IEEE Access*, 8, 32816-32825. <https://doi.org/10.1109/access.2020.2973923>
- Shi, D.; Wang, T.; Xing, H. & Xu, H. (2020). A learning path recommendation model based on a multidimensional knowledge graph framework for e-learning. *Knowledge-Based Systems*. <https://doi.org/10.1016/j.knsys.2020.105618>
- Shin, H. S. & Jeong, A. (2021). Modeling the relationship between students' prior knowledge, causal reasoning processes, and quality of causal maps. *Computers & Education*, 163. <https://doi.org/10.1016/j.compedu.2020.104113>
- Wang, C.; Ma, X.; Chen, J. & Chen, J. (2018). Information extraction and knowledge graph construction from geoscience literature. *Computers & Geosciences*, 112, 112-120. <https://doi.org/10.1016/j.cageo.2017.12.007>
- Wang, H.; Zhang, F.; Wang, J.; Zhao, M.; Li, W.; Xie, X. & Guo, M. (2019). Exploring high-order user preference on the knowledge graph for recommender systems. *ACM Transactions on Information Systems*, 37(3), 1-26. <https://doi.org/10.1145/3312738>
- Wang, Q.; Ding, G. & Yu, S. (2019). Crowdsourcing mode-based learning activity flow approach to promote subject ontology generation and evolution in learning. *Interactive Learning Environments*, 27(7), 965-983. <https://doi.org/10.1080/10494820.2018.1509875>
- Wang, X.; He, X.; Cao, Y.; Liu, M. & Chua, T.-S. (2019). KGAT: Knowledge graph attention network for recommendation.

- KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.
- Wang, Z.; Li, L.; Li, Q. & Zeng, D. (2019). Multimodal data enhanced representation learning for knowledge graphs. 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary.
- Yang, J.; Wang, G.; Zhang, Q. & Wang, H. (2020). Knowledge distance measure for the multigranularity rough approximations of a fuzzy concept. *IEEE Transactions on Fuzzy Systems*, 28(4), 706-717. <https://doi.org/10.1109/tfuzz.2019.2914622>
- Yoo, S. & Jeong, O. (2020). Automating the expansion of a knowledge graph. *Expert Systems with Applications*, 141. <https://doi.org/10.1016/j.eswa.2019.112965>
- Yu, T.; Li, J.; Yu, Q.; Tian, Y.; Shun, X.; Xu, L.; Zhu, L. & Gao, H. (2017). Knowledge graph for TCM health preservation: Design, construction, and applications. *Artificial Intelligence in Medicine*, 77, 48-52. <https://doi.org/10.1016/j.artmed.2017.04.001>
- Zárate, M.; Rosales, P.; Braun, G.; Lewis, M.; Fillottrani, P. R. & Delrieux, C. (2019). OceanGraph: some initial steps toward a oceanographic knowledge graph. *Communications in Computer and Information Science*. KGSWC: Iberoamerican Knowledge Graphs and Semantic Web Conference, Villa Clara, Cuba.
- Zhang, R.; Mao, Y. & Zhao, W. (2020). Knowledge graphs completion via probabilistic reasoning. *Information Sciences*, 521, 144-159. <https://doi.org/10.1016/j.ins.2020.02.016>
- Zhao, M.; Wang, H.; Guo, J.; Liu, D.; Xie, C.; Liu, Q. & Cheng, Z. (2019). Construction of an industrial knowledge graph for unstructured chinese text learning. *Applied Sciences*, 9(13). <https://doi.org/10.3390/app9132720>
- Zhu, Q.; Zhou, X.; Zhang, P. & Shi, Y. (2019). A neural translating general hyperplane for knowledge graph embedding. *Journal of Computational Science*, 30, 108-117. <https://doi.org/10.1016/j.jocs.2018.11.004>



Este artículo es de acceso abierto. Los usuarios pueden leer, descargar, distribuir, imprimir y enlazar al texto completo, siempre y cuando sea sin fines de lucro y se cite la fuente.

CÓMO CITAR ESTE ARTÍCULO:

Hernández-Almazán, Jorge Arturo; Lumbreras-Vega, Juan Diego; Amaya Amaya, Arturo y Machucho-Cadena, Rubén. (2021). Grafo de conocimiento para determinar el dominio del aprendizaje en la educación superior. *Apertura*, 13(1), pp. 118-133. <http://dx.doi.org/10.32870/Ap.v13n1.1937>