

# Representación y aprendizaje de conceptos en Twitter: un análisis de tuits como huellas digitales

## Representation and Learning of Concepts on Twitter: An Analysis of Tweets as Digital Footprints



Mauricio Buitrago-Ropero - *Universidad Libre (Colombia)*

Andrés Chiappe Laverde - *Universidad de La Sabana (Colombia)*

### RESUMEN

Los tuits o mensajes publicados en la red social Twitter son entendidos como huellas digitales que se producen por la interacción de las personas en entornos digitales. Estas huellas se generan tanto en los procesos de educación formal como los que se conducen a través de Ambientes Virtuales de Aprendizaje, como en los procesos de interacción social propios de las redes sociales. En este estudio se analizaron los procesos de representación y aprendizaje de conceptos a partir de la producción de tuits generados por tres grupos de estudiantes universitarios de tercer y quinto año. Dicho estudio de carácter mixto se adelantó bajo el enfoque del aprendizaje supervisado que contempló dos momentos: uno de instrucción y otro de evaluación. Los tuits se analizaron desde tres categorías: contenido, contenedor y contexto, así como desde las operaciones intelectuales del pensamiento conceptual: supra-ordinación, exclusión, infra-ordinación e iso-ordinación. Adicionalmente, se analizó el tono emocional de los tuits mediante técnicas de análisis de contenido, minería de textos y de análisis de sentimiento. Los resultados del estudio señalan la posibilidad de que las huellas digitales puedan ser utilizadas como indicadores de los procesos de representación y aprendizaje de conceptos, no solo desde la perspectiva de la construcción lingüística y cognitiva que supone aprender y representar conceptos, sino desde las condiciones emocionales que se dan en las interacciones de una red social como Twitter. A partir de allí se discuten y se abordan conclusiones relacionadas con el potencial transformador del uso de huellas digitales en educación.

**Palabras clave:** análisis de contenido; aprendizaje; conocimiento; tecnologías de la información y de la comunicación.

### ABSTRACT

The Tweets or messages published on the social network Twitter are understood as digital footprints that are produced by the interaction of people in digital environments. These footprints are generated both in formal education processes such as those conducted through Virtual Learning Environments, as well as in social interaction processes, typical of social media. In this study, the processes of concept representation and learning based on the production of tweets generated by three groups of third and fifth year university students were analyzed. This study of a mixed nature was carried out under the supervised learning approach that included two moments: one of instruction and the other of evaluation. The tweets were analyzed from three categories: content, container and context, as well as from the intellectual operations of conceptual thinking: supra-ordination, exclusion, under-ordination and iso-ordination. Additionally, the emotional tone of the tweets was analyzed using content analysis, text mining and sentiment analysis techniques. The results of the study indicate the possibility that fingerprints can be used as indicators of the processes of representation and learning of concepts, not only from the perspective of the linguistic and cognitive construction involved in learning and representing concepts, but also from the emotional conditions that occur through the interactions within a social network like Twitter. From there, conclusions related to the transformative potential of the use of fingerprints in education are discussed and addressed.

**Keywords:** content analysis; learning; knowledge; information and communication technologies.

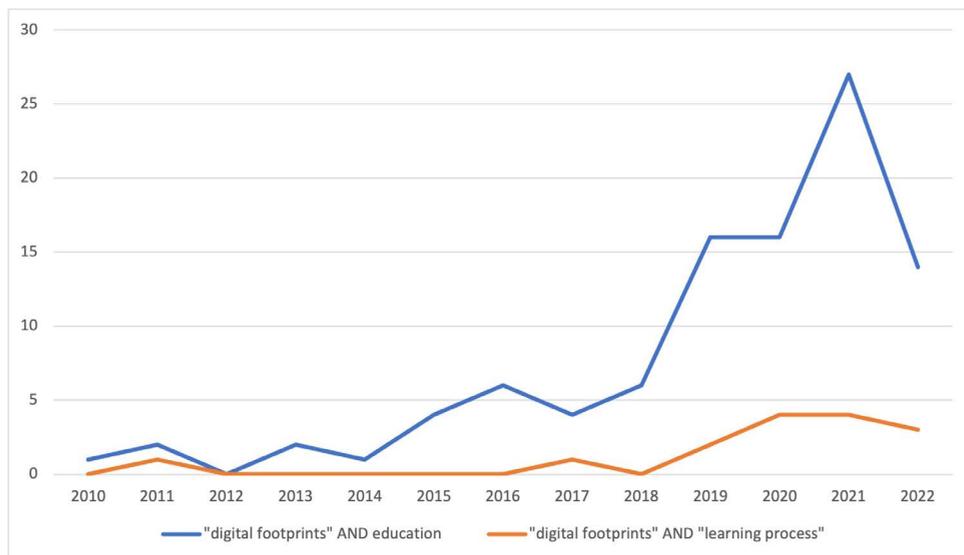
## INTRODUCCIÓN

Las huellas digitales (HD) constituyen el conjunto de datos que las personas producen voluntaria o involuntariamente en sus interacciones digitales a partir del uso de redes sociales (Mori y Haruno, 2021; Wang et al., 2020), portales de comercio electrónico (Khajehasani et al., 2020) mobile\nphones, PDA, digital TVs etc. The number of Internet users is constantly growing, and internet communications\nhave become routine. The use of internet by companies to provide their products and brand is commonplace\nand even inevitable. This paper is focused on marketing by high-speed and low-cost internet. Emarketing\nrefers to all efforts in terms of adjusting and developing marketing strategies in virtual spaces including,\nweb, social media etc. Big and powerful internet marketing websites do not make small internet\nwebsites disappear but enable them to gain competitive advantage by market segmentation strategy. When a\ncustomer visits a website (websites, plataformas de aprendizaje (Chaabi et al., 2020), o al conjunto de electrodomésticos con capacidad de conectarse a la web los cuales hacen parte del llamado “Internet de las cosas” (Mohamed et al., 2022).

En ese contexto, el estudio y uso de las HD con el objetivo de incidir en los procesos de toma de decisiones de las personas es ampliamente utilizado en el sector comercial y de servicios; sin embargo, en el campo de la educación apenas empieza a ser explorado en asuntos como el modelado psicométrico de los estudiantes, el análisis de la retención y la deserción del estudiantado; y, la predicción del éxito, el fracaso y el compromiso escolar, entre otros.

Tal como se muestra en la Figura 1, hay investigaciones sobre el uso de HD en educación pero pocas veces se enfocan en el aprendizaje de conceptos a pesar de ser un tema central en los estudios sobre procesos de construcción y representación del conocimiento (Bin et al., 2020; Chen et al., 2020).

**Figura 1**  
 Investigación sobre huellas digitales en educación



Fuente: Scopus.

Estas huellas, además de ser indicadores de los contenidos que consumen y producen las personas en el mundo digital, también son indicadores de las acciones que realizan y de los servicios que consumen como usuarios de Internet. Lo anterior ha llevado a que las HD sean entendidas de diversas maneras, ya sean como dato (Madden et al., 2007), como acción (Pozdeeva et al., 2021; Sjöberg et al., 2017) o como servicio (Loutfi, 2022), lo que para Buitrago-Ropero et al. (2020) las hace especialmente atractivas en el campo de la educación pues pueden ser estudiadas desde su dimensión teórica, práctica y axiológica.

Por otra parte, Cobo (2016) señala que el desarrollo de procesos educativos mediados por tecnologías digitales supone pensar en tres categorías: el contenido, el contenedor y el contexto. El Contenido se refiere al conjunto de saberes que constituyen los distintos corpus teóricos que se pueden enseñar de una disciplina científica. El aprendizaje de estos corpus configura redes complejas de conceptos las cuales representan el conocimiento conceptual de una persona. El Contenedor está referido al medio que actúa como vehículo para que los contenidos puedan ser almacenados, sistematizados y comunicados en el desarrollo de las actividades de aprendizaje. En el mundo digital, el vehículo o contenedor es Internet y las HD configuran rastros tanto de los datos que producen/consumen las personas como de las acciones y servicios que utilizan. Por último, el Contexto debe entenderse como

el conjunto de condiciones que rodean una práctica o acción educativa. Cobo (2016, p. 63) indica que estos contextos “están influenciados por una vasta cantidad de factores, por ejemplo, institucionales, normativos, relacionales, sociales, culturales, políticos, económicos, emocionales, entre otros”.

### **Primer acercamiento: representar el conocimiento**

Según Ouyang et al. (2023), el conocimiento se declara a partir de redes de oraciones que sirven como indicadores de la actividad cognitiva. Este conocimiento implica la construcción de redes de conceptos que se interconectan y declaran a través de proposiciones (Murnikov y Kask, 2021) que se producen en el marco de cuatro operaciones intelectuales del pensamiento conceptual (OIPC), estas son: la supra-ordinación, la exclusión, la infra-ordinación y la iso-ordinación. La supra-ordinación se genera cuando se determina la categoría conceptual que contiene y subordina el concepto, así la supra-ordinación del concepto perro es caninos. La exclusión se produce cuando se enuncian las categorías conceptuales análogas al concepto las cuales comparten con éste la misma supra-ordinación; por ejemplo, las exclusiones del concepto perro son lobo o zorro. Entre tanto, las infra-ordinaciones corresponden a las categorías conceptuales a partir de las cuales se puede subdividir el concepto. Algunas infra-ordinaciones del concepto perro serían entonces labrador retriever (raza) o perro de compañía (usos). Por último, la iso-ordinación se construye cuando se declaran los rasgos esenciales y distintivos del concepto frente a sus exclusiones. Así, una iso-ordinación de perro estaría relacionado con su condición de ser el único canino domesticado.

### **Segundo acercamiento: las redes sociales como contenedores de huellas digitales**

De acuerdo con las cifras del Pew Internet Research Center, se estima que las redes sociales más usadas son: Youtube (n=73 %), Facebook (n=68 %), Instagram (37 %), Pinterest (28 %), LinkedIn (27 %), Snapchat (n=24 %), Twitter (n=22 %), WhatsApp (n=20 %) y ReddIt (n=11 %). Así, las redes sociales, se siguen extendiendo como medios de producción de huellas digitales y su uso es generalizado pues hacen parte de la cotidianidad de las personas (Argente et al., 2017). Al respecto, Sheikh et al. (2021) resaltan el crecimiento constante de las redes sociales como un fenómeno que se proyecta por encima de los 3430 millones de usuarios para el año 2023, llegando a representar más del 82 % de la totalidad de los internautas de social media a nivel global. Este fenómeno ha hecho que las redes sociales se configuren como los contenedores más usados en el mundo digital.

En particular, los tuits como huellas digitales y como expresiones lingüísticas digitales, tienen una serie de características propias de lo que autores como Nixon y Guajardo (2022) denominan textismos. Allí adquieren la forma de escritos no

convencionales que se apoyan además en gráficos digitales como los llamados emoticones y en expresiones abreviadas que parecieran ser la norma en la escritura digital. Como se sabe, esta escritura digital, al menos en el contexto de las redes sociales está caracterizada por no cumplir varias normas gramaticales y de escritura académica. De esta forma, las HD provenientes de estos entornos digitales han de estudiarse bajo este tipo de consideraciones que las ubica como textos digitales con elementos lexicales y semánticos nuevos que pueden ayudar a entender la forma como percibimos e interactuamos con los demás (Nixon y Guajardo, 2022).

### **Tercer acercamiento: el tono emocional como elemento de análisis del contexto de las huellas digitales**

Estudios como los de Buitrago-Ropero et al. (2020) señalan que el análisis de la emocionalidad que rodea el proceso de construcción de conocimiento está ubicado en el tercer lugar de los estudios sobre el uso de HD en educación. En particular, estos estudios se utilizan para la determinación y el análisis de rasgos de personalidad, tono emocional y tono del lenguaje, entre otros. Para ello, se usa la minería de textos la cual recurre a los principios básicos de la lingüística computacional con propósitos que van desde el etiquetado gramatical, hasta el análisis de estados emocionales a partir de señales psicolingüísticas tal como lo describen Li et al. (2022). Entre tanto, el estudio del tono emocional suele vincular variables como el sentimiento y la ironía mientras que el tono de lenguaje se asocia a variables como la polaridad, la confianza y la objetividad textual.

## **MÉTODO**

El estudio se adelantó con tres grupos de estudiantes universitarios quienes participaron de un proceso de entrenamiento y evaluación de conceptos que debían ser explicados en Twitter a lo largo de un periodo de 15 semanas (ver Tabla 1). El análisis de los tuits se realizó a partir de los elementos que señala Cobo (2016) en su Teoría de las 3C: (1), Contenido, (2), Contenedor y (3) Contexto, correspondientes a los siguientes elementos: (1) los tuits que produjeron los estudiantes como reacción a preguntas del docente los cuales fueron denominados tuits conceptuales (TC), (2) la red social Twitter; y por último, (3) el tono emocional y de lenguaje desde donde se produjeron los tuits.

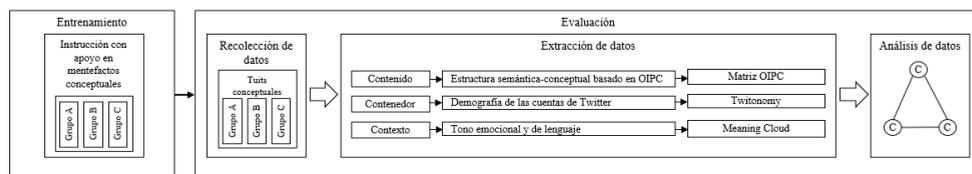
**Tabla 1**  
*Participantes del estudio*

Grupo	Clase	Año	Estudiantes	%
Gr. A	Informática Aplicada	Tercero	23	31.94
Gr. B	Informática Aplicada	Tercero	37	51.39
Gr. C	Tecnologías de la Información y la Comunicación (TIC)	Quinto	12	16.67
Total			72	100

## Fases de la investigación

Tal como muestra la Figura 2, el estudio se desarrolló en 2 fases y siguió el enfoque del aprendizaje supervisado (Rajesh Kumar et al., 2020), el cual contempla dos momentos: entrenamiento y evaluación.

**Figura 2**  
*Fases de la investigación*



*Fuente:* Elaboración propia basada en las fases del Aprendizaje Supervisado.

## Fase de Entrenamiento

Esta fase se destinó al proceso de instrucción y enseñanza de conceptos relacionados con los temas abordados en las asignaturas que cursaron los tres grupos de estudiantes participantes.

Como parte de la estrategia de enseñanza de conceptos, se utilizaron los mentefactos conceptuales y se explicó cómo, a partir de estos, se pueden determinar las cuatro categorías que circunscriben un concepto; estas categorías se concretan en las cuatro OIPC.

## Fase de Evaluación

Esta fase se extendió por 15 semanas y se dividió en tres procesos orientados a la recolección, la extracción y el análisis de los tuits.

### La Recolección de Datos

El primer momento se destinó a la generación y recolección de los tuits para lo cual los estudiantes ( $n=72$ ) reaccionaron voluntariamente a preguntas planteadas por el docente. En los tuits se indagó por los conceptos enseñados en la fase de entrenamiento. Para esto se creó una cuenta en Twitter bajo el nombre de @HuellasDigital3. Los tuits de respuesta fueron denominados tuits conceptuales (TC). De esta forma, el set de datos sobre el que se hizo el proceso de extracción, sistematización y análisis de las HD se configuró tal como se muestra en la Tabla 2. Los tuits respondieron a preguntas sobre los conceptos enseñados en las clases y no incluyeron el uso de emoticones (aunque no se planteó como requisito) pero sí de expresiones abreviadas o que no seguían algunas reglas gramaticales configurando así algunos textismos.

**Tabla 2**

*Variables y valores del conjunto de datos*

Ítem	Cantidad	Ítem	Cantidad
Estudiantes tuiteros (ET)	72	Total caracteres en los TC	78434
Semanas de actividad de tuiteo conceptual	15	Prom. Palabras x TC	35.71
Grupos de clase (A, B y C)	3	Prom. Caracteres x TC	222.82
Total tuits conceptuales (TC)	352	Prom. Estud. Con reacción a los TC	23.47
Total tuits no conceptuales (TNC)	96521	% Promedio de participación de los ET en los TC	32.59%
Total tuits	96873	Prom. Caracteres en los TC x estudiante	1089.36
Total palabras en los TC	12571	% de uso de la cadena de caracteres máx. por tuit (280)	79.58%

## La Extracción de Datos

El segundo momento correspondió al proceso de extracción de tuits escritos por los estudiantes. Para ello, se publicaron tuits en la cuenta @HuellasDigital3 a lo largo de 15 semanas (con una tasa de 1 tuit por semana) en las que se indagó por la comprensión de igual número de conceptos (base de datos, exhaustividad, precisión, ruido, silencio, etc.). Para cada pregunta, los estudiantes, de forma voluntaria, explicaron qué comprendían en relación con los conceptos enseñados previamente en clase. Lo anterior, tal como se muestra en la Tabla 3, implicó una participación de los estudiantes con un promedio de reacción de 23,47 estudiantes por tuit.

**Tabla 3**  
*Cantidad de estudiantes con reacción a los TC*

TC	# de Estudiantes con reacción a los TC	% de Participación	TC	# de Estudiantes con reacción a los TC	% de Participación
1	20	27,78	9	33	45,83
2	25	34,72	10	29	40,28
3	9	12,50	11	26	36,11
4	12	16,67	12	18	25,00
5	35	48,61	13	19	26,39
6	35	48,61	14	15	20,83
7	36	50,00	15	6	8,33
8	34	47,22	Promedio	23,47	32,59

Los tuits fueron sistematizados a la luz de tres categorías: Contenido, Contenedor y Contexto.

### Categoría 1: Contenido

Para el proceso de extracción de tuits desde la categoría Contenido, se utilizó la estrategia de la estructuración semántica conceptual basada en la construcción de mentefactos conceptuales y en la determinación de la relación entre los TC y las 4 OIPC. Estas operaciones debieron evaluarse a partir de una estructura previa de mentefactos elaborados por tres expertos temáticos que explicaron, a través de estos organizadores gráficos, los 15 conceptos abordados en clase. La validez de cada mentefacto se realizó mediante el coeficiente de Kappa de Cohen para el cual se

estableció como valor mínimo de referencia  $K=0,81$  correspondiente a un valor muy alto de concordancia según lo indican Domene-Martos et al. (2021).

De esta forma, cada TC producido por los estudiantes ( $n=72$ ) fue entendido como un conjunto de Oraciones Proposicionales (OP) que daban cuenta de la comprensión de un concepto. Estas OP se fueron registrando en una matriz en la que se indicaba si ellas reflejaban una supra-ordinación, una exclusión, una infra-ordinación y/o una iso-ordinación. Adicionalmente, se fue señalando con un código de colores (véase a modo de ejemplo la Tabla 4) si las OP eran correctas (verde), parcialmente correctas (amarillo) o incorrectas (rojo).

**Tabla 4**

*Extracción y sistematización de los TC a partir del análisis de OIPC*

Contenido del TC	Supra-ordinación	Exclusión	Infra-ordinación	Iso-ordinación
Las bases de datos académicas son sistemas de información especializados. Algunos ejemplos son Scopus, Redalyc e ISI Web of Science también conocido como WoS	X (verde)		X (verde)	
El ruido, al igual que el silencio, es uno de los fenómenos presentes en todo proceso de búsqueda de información	X (verde)	X (verde)		
Los ambientes virtuales de aprendizaje hacen parte de los ambientes de aprendizaje presenciales y su rasgo principal es que se apoyan en el uso de TIC. Se clasifican según su entorno y estos pueden ser de conocimiento, colaboración, asesoría experimentación o de gestión.	X (rojo)		X (amarillo)	X (verde)

## Categoría 2: Contenedor

Para el proceso de extracción de tuits desde la categoría Contenedor, se utilizaron tanto los TC como los tuits que producían los estudiantes en su actividad normal en Twitter (tuits no conceptuales o TNC). La estrategia consistió en el estudio demográfico del comportamiento de la cuenta de cada estudiante en el periodo de tiempo en el que se llevó a cabo la generación de TC. Para ello, se utilizó el programa en línea Twitonomy y se analizaron las siguientes variables: total de tuits generados,

porcentaje de respuesta a otros tuits (retuits), número de seguidores (N), número de personas seguidas (M), día de mayor actividad, hora de mayor actividad y plataforma de tuiteo.

### Categoría 3: Contexto

Para el proceso de extracción y sistematización de tuits desde la categoría Contexto, se utilizaron nuevamente los TC y la estrategia del análisis del tono emocional y de lenguaje a partir de 5 variables: sentimiento, confianza, polaridad, ironía y objetividad textual. Cada una de estas variables se midió en términos de los indicadores señalados en la Tabla 5. El estudio de los tonos se llevó a cabo con la API Sentiment Analysis que hace parte del programa en línea Meaning Cloud.

**Tabla 5**

*Variables e indicadores para el análisis del tono emocional*

Variable	Indicador
Sentimiento	Muy positivo, positivo, negativo, muy negativo, neutral, sin sentimiento
Confianza	Escala numérica (0-100)
Polaridad	Acuerdo, desacuerdo
Ironía	Con ironía, sin ironía
Textualidad	Objetiva, subjetiva

## RESULTADOS

Los resultados se presentan en función de las tres categorías propuestas en la metodología descrita en el apartado anterior: Contenido, Contenedor y Contexto.

### Resultados de la categoría Contenido

Bajo esta categoría, se tomaron los tuits que produjeron los estudiantes como reacción a las preguntas que el docente publicó con el propósito de indagar por algunos de los conceptos abordados en clase. A estos tuits, se les denominó tuits conceptuales (TC). Para cada TC, el número de estudiantes que reaccionó fue distinto por cuanto la participación de ellos se realizó de forma voluntaria durante toda la investigación. Al respecto, el detalle de la participación de los estudiantes se muestra en la Tabla 6.

**Tabla 6***Detalle de la participación de los estudiantes en Twitter*

TC	# Estudiantes con Reacción al TC	Cantidad de OP	Tasa OP/EST
1	20	53	2,65
2	25	35	1,40
3	9	14	1,56
4	12	30	2,50
5	35	68	1,94
6	35	79	2,26
7	36	68	1,89
8	34	62	1,82
9	33	65	1,97
10	29	58	2,00
11	26	34	1,31
12	18	43	2,39
13	19	25	1,32
14	15	22	1,47
15	6	15	2,50
Prom.	23,46	44,73	1,93

Si se cruzan los datos de la Tabla 6 con los de la Tabla 3, se observa entonces que los estudiantes escriben cerca de dos oraciones proposicionales en cada TC usando el 79,58 % de la capacidad de caracteres máximo permitido en Twitter (n=280), es decir, un promedio de 222,82 caracteres por tuit.

Los tuits fueron dispuestos en una matriz en la que se analizaron las oraciones proposicionales inmersas en cada tuit con el ánimo de identificar cuáles de las 4 OIPC se encontraban representadas en dichas oraciones (Tabla 7).

**Tabla 7***Cantidad de OP y tipos de OIPC involucradas en los TC*

TC	Cantidad de OP	Tipo de OIPC involucradas				
		S	E	IN	IS	ND
1	53	16	1	6	15	15
2	35	0	0	5	5	25
3	14	5	0	0	9	0

TC	Cantidad de OP	Tipo de OIPC involucradas				
		S	E	IN	IS	ND
4	30	7	3	1	10	9
5	68	19	0	16	29	4
6	79	30	3	24	20	2
7	68	25	4	11	27	1
8	62	23	10	3	26	0
9	65	26	4	4	29	2
10	58	29	0	2	27	0
11	34	9	0	0	25	0
12	43	12	4	0	15	12
13	25	0	0	4	2	19
14	22	0	0	10	12	0
15	15	6	0	1	4	4
Total	671	207	29	87	255	93

Algunas de las oraciones no se catalogaron bajo ninguna OIPC por lo cual se les designó como no determinadas (ND). Por otra parte, las oraciones proposicionales (OP) fueron evaluadas a partir de tres escalas: correcta, incorrecta y parcialmente correcta; esto para reconocer si el ejercicio de conceptualización que se mostraba en los TC era elaborado adecuadamente. Los resultados se muestran en la Tabla 8.

**Tabla 8**

*Evaluación de los TC en función de las OIPC desplegadas en cada TC*

OIPC desplegadas en los TC	Valoración	Cantidad de OP	%	Cantidad total de OP x OIPC	%
Supra-ordinación	Correcta	65	9,69	207	30,85
	Incorrecta	64	9,54		
	Parcialmente correcta	78	11,62		
Exclusión	Correcta	2	0,30	29	4,32
	Incorrecta	6	0,89		
	Parcialmente correcta	21	3,13		
Infra-ordinación	Correcta	16	2,38	87	12,97
	Incorrecta	8	1,19		
	Parcialmente correcta	63	9,39		

OIPC desplegadas en los TC	Valoración	Cantidad de OP	%	Cantidad total de OP x OIPC	%
Iso-ordinación	Correcta	73	10,88	255	38
	Incorrecta	70	10,43		
	Parcialmente correcta	112	16,69		
No definidas	N/A	93	13,86	93	13,86
Total				671	100

En total, los estudiantes produjeron 671 OP equivalentes a 12571 palabras y 78434 caracteres. De éstas, la mayor actividad se concentró en oraciones del tipo iso-ordinación (n=255) y la menor actividad se presentó en oraciones del tipo exclusión (n=29). La OIPC donde mayor nivel de acierto hubo fue la iso-ordinación (n=10.88%) que a su vez corresponde con el tipo de oraciones que más produjeron los estudiantes (n=38 %) en su actividad de tuiteo conceptual.

En relación con los grupos de estudiantes participantes (Gr. A, Gr. B y Gr.C), los resultados muestran que los estudiantes más activos en la producción de TC correspondieron al Grupo B. Este grupo, además de ser el más numeroso, es el que mayor nivel de representatividad tiene sobre la producción de TC. El detalle de la participación por grupos se muestra en la Tabla 9.

**Tabla 9**

*Distribución de estudiantes, TC y tasa de participación por grupo de clase*

Grupo	Clase	Estudiantes	%	Cantidad de TC	%	Tasa de participación
Gr. A	Informática Aplicada	23	31,94	74	21,02	0.65
Gr. B	Informática Aplicada	37	51,39	231	65,63	1.27
Gr. C	NTIC	12	16,67	47	13,35	0.80
Total		72	100,00	352	100,00	NA

## Resultados de la categoría Contenedor

Bajo esta categoría, se analizó el comportamiento de la cuenta de cada estudiante en Twitter. El análisis consideró tanto la actividad de producción de tuits conceptuales (TC) como de los tuits que produjeron los estudiantes como parte de su actividad habitual en esta red social (TNC).

La herramienta de análisis escogida fue el programa Twitonomy el cual permite hacer una radiografía amplia de la cuenta de Twitter de una persona mostrando datos como el número de seguidores (N), número de personas que se sigue (M), el crecimiento diario de la cuenta y el promedio de tuits producidos diariamente. Así, el estudio buscó analizar datos como: día de la semana de mayor actividad, hora de mayor actividad, plataforma de mayor uso para tuitear, N, M y la Tasa N/M; esta última fue usada como indicador del nivel de participación en la red. Se analizaron las cuentas del 100 % de los estudiantes que participaron en la investigación y se estudiaron las variables relacionadas con la categoría Contenedor, cuyo detalle se muestra en la Tabla 10.

**Tabla 10**  
*Variables analizadas para la categoría Contenedor*

<b>Variables</b>	<b>Valor</b>
Total Tuits (TC + TNC)	96873
Total Retuits	1019
Valor máximo de tuits de un estudiante	24198
Valor mínimo de tuits de un estudiante	1
Promedio Tuits por día	0,42
Promedio N	81,26
Promedio M	112,99
Tasa Promedio N/M	0,83

Por otra parte, el análisis de la actividad digital de los participantes se realizó en función del día de la semana y las horas del día en que se produjeron los tuits (Tablas 11 y 12). Esta relación revelaría asuntos como la gestión del tiempo de estudio y de desarrollo de actividades de aprendizaje orientadas a la apropiación conceptual.

**Tabla 11**  
*Análisis de la actividad de tuiteo en función del día de la semana*

<b>Día</b>	<b>%</b>
Lunes	34,72
Martes	15,28
Miércoles	11,11
Jueves	8,33
Viernes	5,56
Sábado	6,94

Día	%
Domingo	15,28
No disponible	2,78
Total	100,00

**Tabla 12**

*Análisis de la actividad de tuiteo en función de la hora del día*

Hora del día	%	Hora del día	%
1:00-1:59	15,28	13:00-13:59	8,33
2:00-2:59	19,44	14:00-14:59	4,17
3:00-3:59	6,94	15:00-15:59	2,78
4:00-4:59	5,56	16:00-16:59	2,78
5:00-5:59	1,39	17:00-17:59	2,78
6:00-6:59	0,00	18:00-18:59	1,39
7:00-7:59	1,39	19:00-19:59	0,00
8:00-8:59	0,00	20:00-20:59	1,39
9:00-9:59	0,00	21:00-21:59	5,56
10:00-10:59	0,00	22:00-22:59	6,94
11:00-11:59	0,00	23:00-23:59	6,94
12:00-12:59	0,00	No disponible	2,78
Total			100,00

Por último, se muestra el análisis comparativo de la actividad de tuiteo para los 3 grupos (Gr. A, Gr. B y Gr.C). Los datos de la Tabla 13 reflejan que el grupo que más intervino en la producción de tuits fue el Grupo B con más del 50 % de la participación. En relación con su condición de usuarios que siguen o son seguidos por otros, el grupo que mejor participación tuvo fue el Grupo C (N/M=1,44).

**Tabla 13**

*Análisis comparativo de la actividad de tuiteo por grupos*

Grupo	Est	%	Total tuits	%	TNC	%	TC	%	Día mayor actividad	Hora mayor actividad	N/M	Plataforma
Gr. A	23	31,94	13699	14,14	13625	14,12	74	21,02	Domingo	2:00 a.m.	0,43	Twitter Web (43,06%)
Gr. B	37	51,39	29772	30,73	29541	30,61	231	65,63	Domingo	2:00 a.m.	0,66	Android
Gr. C	12	16,67	53402	55,13	53355	55,28	47	13,35	Miércoles	2:00 a.m.	1,44	(41,67%)

### Resultados de la categoría Contexto

Bajo esta categoría, se analizó el tono emocional y de lenguaje de los TC de los estudiantes. La herramienta de análisis fue Sentiment Analysis, una interfaz de programación de aplicaciones semánticas (API semántica). Sentiment Analysis permite estudiar variables como el sentimiento, la confianza, la polaridad, la textualidad y la ironía asociadas a la producción lingüística de una persona. Para el análisis, las palabras que conformaban cada TC fueron tratadas como unidades psicolingüísticas usando el enfoque del aprendizaje automático (Mir et al., 2022). La descripción analítica de estas variables se presenta en la Tabla 14.

**Tabla 14**

*Descripción de las variables usadas para analizar el tono emocional y de lenguaje*

Tono	Variable	Descripción
Emocional	Sentimiento	Analiza la carga afectiva global que se produce en la manifestación de una idea o concepto.
	Ironía	Analiza la existencia de expresiones que contradicen una idea o concepto. Junto con la polaridad, sirve para medir la contradicción interna de una idea o concepto.

Tono	Variable	Descripción
De lenguaje	Confianza	Analiza el nivel de certeza con el que se afirma una idea o concepto. Su medida es usualmente cuantitativa indicando o el valor más bajo de certeza y 100 el más alto.
	Polaridad	Analiza una expresión en función de lo contradictorio que pueden llegar a resultar las distintas partes que se usan en un texto para afirmar o negar algo. Es una medida de la contradicción interna en la manifestación de una idea o concepto.
	Objetividad textual	Analiza el nivel de construcción analítica usualmente ligada a condiciones como la racionalidad, la sistematicidad y la inclusión o no de juicios de valor individual expresados en una idea o concepto.

En la investigación, se llevó a cabo el análisis del 100 % de los TC (n=352) que produjeron los estudiantes. Como se observa en la Tabla 15, la generación de TC tuvo tendencia a producirse desde sentimientos positivos (n>69 %), con una polaridad predominantemente valorada como de acuerdo; por otra parte, la textualidad fue calificada principalmente como objetiva (n=52 %) aunque sin distar mucho de su opuesta, la textualidad subjetiva. Todo lo anterior se produjo en un contexto sin ironía manifiesta. Entre tanto, el promedio de confianza se ubicó en un rango muy alto (n>97) con valores en todos los TC por encima de 75.

**Tabla 15**  
*Análisis del tono emocional y de lenguaje de los TC*

Variable	Indicador	Cantidad de TC	%
Sentimiento	Muy positivo	51	14
	Positivo	198	56
	Negativo	51	14
	Muy negativo	5	1
	Neutral	19	5
	Sin sentimiento	28	8
	<b>Total</b>	<b>352</b>	<b>100</b>
Ironía	Con ironía	0	0
	Sin ironía	352	100
	<b>Total</b>	<b>352</b>	<b>100</b>

Variable	Indicador	Cantidad de TC	%
Polaridad	Acuerdo	281	80
	Desacuerdo	71	20
	<b>Total</b>	<b>352</b>	<b>100</b>
Objetividad textual	Objetivo	183	52
	Subjetivo	169	48
	<b>Total</b>	<b>352</b>	<b>100</b>
Confianza Promedio		97,41	
Valor máximo confianza		100	
Valor mínimo confianza		76	

En relación con los grupos, se observa en la Tabla 16 una fuerte influencia del Grupo B en la determinación de la variable Sentimiento, ubicándose como grupo más destacado en al menos tres indicadores, especialmente los que determinan la condición positiva. Entre tanto, para esta misma variable, el Grupo C se presenta como el de menor relevancia en 4 de los indicadores. Para la variable Polaridad, el grupo que mayor nivel de polaridad positiva presentó fue nuevamente el Grupo B mientras que el Grupo A se muestra como el de mayor polaridad negativa. En términos de la Objetividad textual, el Grupo B se destaca por ser el que moviliza el conjunto de TC hacia la objetividad, mientras que el Grupo C lo hace hacia la subjetividad. El nivel de Ironía detectado en todos los grupos es nulo.

**Tabla 16**  
*Preponderancia de los grupos de acuerdo con las variables del tono emocional*

Variable	Indicador	Grupo más preponderante	Grupo menos preponderante
Sentimiento	Muy positivo	B	C
	Positivo	B	C
	Negativo	B	C
	Muy negativo	A	C
	Neutral	A	B
	Sin sentimiento	B	A
Ironía	Con	Ninguno	Ninguno
	Sin	A, B y C	A, B y C
Polaridad	Acuerdo	B	A
	Desacuerdo	A	B
Objetividad textual	Objetiva	B	C
	Subjetiva	C	B

## DISCUSIÓN Y CONCLUSIONES

Abordar las huellas digitales como indicadores de los procesos de representación del conocimiento en entornos de aprendizaje digital abre la posibilidad de entenderlas más allá de los datos mismos. Su análisis debe hacerse desde un panorama amplio en el que se estudie su relación con los dispositivos y recursos tecnológicos que se utilizan para su difusión, así como con las condiciones emocionales y de lenguaje que subyacen a su producción. Lo anterior, se convierte en información pertinente para el estudio profundo de los procesos de enseñanza y aprendizaje en espacios como los ambientes virtuales de aprendizaje, los cursos masivos abiertos en línea (MOOC) o las redes sociales, tal como lo señalan Chamorro et al. (2022) y Zubareva et al. (2022).

Con relación al uso de operaciones intelectuales del pensamiento conceptual (OIPC) implicadas en la representación del conocimiento conceptual, se infiere una gran dificultad en los estudiantes para categorizar un concepto usando de forma simultánea y correcta las 4 OIPC. El hecho de que los estudiantes hagan un énfasis marcado sobre procesos de iso-ordinación y supra-ordinación puede indicar una mayor capacidad por parte de los estudiantes para establecer los rasgos distinguibles (iso-ordinación) de un concepto en relación con otros conceptos análogos, así como para comprender el concepto como parte de un concepto mayor que lo contiene (supra-ordinación). El vacío o la brecha de aplicación sobre las otras dos OIPC serviría de alerta para ajustar los procesos de enseñanza de conceptos.

De otro parte, llama la atención que pese a la limitante de los 280 caracteres que impone Twitter, los estudiantes usan un promedio de 1,93 oraciones proposicionales y el 79,58% de la capacidad máxima de caracteres. Al respecto, la cantidad de oraciones proposicionales (OP) parece no tener relación con el acierto o el desacierto conceptual. Esto podría indicar que la actividad de comprensión de conceptos y de escritura de OP, más que estar relacionada con la extensión textual, está vinculada, por una parte, con la capacidad de escritura en línea en el nivel más básico (descriptivo) tal como lo explica Cassany (2012); y por otra, con procesos de memorización (Bautista-Vallejo et al., 2020) más que con procesos de desarrollo del pensamiento conceptual y la construcción colaborativa y folksonómica del conocimiento tal como lo afirman Yu y Chen (2020).

Por su parte, el análisis de Twitter como Contenedor de tuits permitió entender que su producción está relacionada con tres condiciones: la instrucción, la asignación de deberes académicos y el uso del tiempo libre asociado al consumo de redes sociales. Al respecto, se encontró que los días de mayor actividad en la producción de tuits (lunes, martes y domingo) siempre estuvieron cerca (o coincidieron) a los días en los que se llevaba a cabo algún proceso de instrucción (martes y miércoles) o de realización de tareas escolares, lo que suele acontecer los días del fin de semana según lo indican los mismos estudiantes. Lo anterior puede ayudar a entender la necesidad de que las prácticas educativas que vinculan el uso de recursos TIC se conecten con

las actividades no necesariamente escolares pues ello puede llevar a un uso más académico de recursos como las redes sociales. Igualmente, los datos confirman que las altas horas de la noche y las primeras horas del día son las preferidas por los estudiantes para llevar a cabo su participación en redes sociales. Este fenómeno se invierte en las horas en las que los estudiantes asisten a clases (7:00 a.m. a 12:00 m. y de 1:00 p.m. a 7:00 p.m.) reflejando el poco o nulo uso de ellas en las actividades educativas que orientan los docentes.

Al mismo tiempo, los indicadores de personas que siguen (N) o son seguidas por los estudiantes (M), reflejan que ellos se perfilan más como productores que como consumidores de HD. Este fenómeno puede resultar útil tanto para el diseño de actividades de aprendizaje colaborativas en entornos digitales que hagan uso de redes sociales como en futuras investigaciones para estudiar el comportamiento prosumidor (de producción y consumo) de HD del que habla Bratianu (2022).

Finalmente, con relación al tono emocional y de lenguaje que subyace a la generación de HD como los tuits, es posible concluir que su caracterización, constituida mayoritariamente por sentimientos catalogados como muy positivos, puede estar asociada al hecho de que la actividad de tuiteo conceptual se planteó desde el inicio como una actividad voluntaria no sujeta a calificación y con una implicación muy baja en términos de dedicación de tiempo. Esta percepción fue ratificada por los estudiantes a lo largo de la fase de evaluación quienes indicaron que la producción de TC fue incorporada de forma natural a su actividad tuitera habitual, especialmente entre los que presentan una alta actividad en Twitter.

En materia de prospectiva de la investigación, se plantea la necesidad de fortalecer el estudio del impacto de la categoría Contexto a partir de enfoques de minería textual basados en el uso de recursos de diccionario cerrado no acotados únicamente a los diccionarios de datos del programa, de modo que los investigadores puedan personalizar y co-construir tales diccionarios (lo cual es hoy técnicamente posible). De otra parte, se hace necesario poder estudiar los tuits como unidades digitales escritas que permiten el uso de textismos y de expresiones gráficas como los emoticones como parte de la composición léxico-semántica de los estudiantes haciendo posible que los estudios analicen dichas expresiones como indicadores de las actividades de comprensión de conceptos, pero también del ámbito emocional desde el que se producen.

Igualmente, se señala la necesidad de que las investigaciones futuras estudien el uso de HD en educación desde la triada Contenido-Contenedor-Contexto en función de otros contenedores digitales tanto de acceso abierto como de acceso cerrado; esto llevaría a estudiar si la representación y la construcción de conceptos expuestos en formatos textuales alfabéticos difiere de la que se hace con los formatos audiovisuales como sucede en YouTube, en donde habría que fijar unas semánticas de análisis distintas.

## Agradecimientos

Agradecemos a la Universidad Libre (Ágora Latinoamericana) y a la Universidad de La Sabana (Grupo Tecnologías para la Academia – Proventus (Proyecto EDUPHD-20-2022) por el apoyo recibido en la preparación de este artículo.

## REFERENCIAS

- Argente, E., Vivancos, E., Alemany, J., y García-Fornes, A. (2017). Educando en privacidad en el uso de las redes sociales. *Education in the Knowledge Society*, 18(2), 107-126. <https://doi.org/10.14201/eks2017182107126>
- Bautista-Vallejo, J. M., Hernández-Carrera, R. M., Moreno-Rodríguez, R., y Lopez-Bastias, J. L. (2020). Improvement of memory and motivation in language learning in primary education through the interactive digital whiteboard (IDW): The future in a post-pandemic period. *Sustainability (Switzerland)*, 12(19). <https://doi.org/10.3390/su12198109>
- Bin, L., Guang, M., Hong, J., y Jigui, Z. (2020). Knowledge Evolution Research on Enterprise Human Resources Management Based on Knowledge Mapping. *Journal of Physics: Conference Series*, 1607, 012113. <https://doi.org/10.1088/1742-6596/1607/1/012113>
- Bratianu, C. (2022). Knowmads as Possible Mutants of Knowledge Workers in the Brave post-COVID World. *Electronic Journal of Knowledge Management*, 20(3), 122-137. <https://doi.org/10.34190/ejkm.20.3.2570>
- Buitrago-Roperero, M. E., Ramírez-Montoya, M. S., y Laverde, A. C. (2020). Digital footprints (2005-2019): A systematic mapping of studies in education. *Interactive Learning Environments*, 1-14. <https://doi.org/10.1080/10494820.2020.1814821>
- Cassany, D. (2012). *En línea. Leer y escribir en la red*. Anagrama.
- Chaabi, Y., Ndiaye, N. M., y Lekdioui, K. (2020). Personalized recommendation of educational resources in a MOOC using a combination of collaborative filtering and semantic content analysis. *International Journal of Scientific and Technology Research*, 9(2), 3243-3248. Scopus.
- Chamorro-Atalaya, O., Arce-Santillan, D., Morales-Romero, G., Ramos-Salazar, P., León-Velarde, C., Auqui-Ramos, E., y Levano-Stella, M. (2022). Sentiment analysis through twitter as a mechanism for assessing university satisfaction. *Indonesian Journal of Electrical Engineering and Computer Science*, 28(1), 430-440. <https://doi.org/10.11591/ijeecs.v28.i1.pp430-440>
- Chen, T., Zhang, S., Wang, Y., Chen, Z., y Jing, W. (2020). Construction Methods of Knowledge Mapping for Full Service Power Data Semantic Search System. *Journal of Signal Processing Systems*. <https://doi.org/10.1007/s11265-020-01591-6>
- Cobo, C. (2016). *La innovación pendiente. Reflexiones (y provocaciones) sobre educación, tecnología y conocimiento*. Penguin Random House.
- Domene-Martos, S., Rodríguez-Gallego, M., Caldevilla-Domínguez, D., y Barrientos-Báez, A. (2021). The use of digital portfolio in higher education before and during the COVID-19 pandemic. *International Journal of Environmental Research and Public Health*, 18(20). <https://doi.org/10.3390/ijerph182010904>
- Khajehasani, S., Abolizadeh, A., y Dehyadegari, L. (2020). The Role

- of Management and Strategy in the Development of E-Marketing. *Recent Advances in Computer Science and Communications*, 13(4), 641-649. <https://doi.org/10.2174/2213275912666190411114639>
- Li, Y., Kazemeini, A., Mehta, Y., y Cambria, E. (2022). Multitask learning for emotion and personality traits detection. *Neurocomputing*, 493, 340-350. <https://doi.org/10.1016/j.neucom.2022.04.049>
- Loutfi, A. A. (2022). A framework for evaluating the business deployability of digital footprint based models for consumer credit. *Journal of Business Research*, 152, 473-486. <https://doi.org/10.1016/j.jbusres.2022.07.057>
- Madden, M., Fox, S., Smith, A., y Vitak, J. (2007, December 16). *Digital Footprints*. Pew Research Center. <https://www.pewresearch.org/internet/2007/12/16/digital-footprints/>
- Mir, A. A., Rathinam, S., y Gul, S. (2022). Public perception of COVID-19 vaccines from the digital footprints left on Twitter: analyzing positive, neutral and negative sentiments of Twitterati. *Library Hi Tech*, 40(2), 340-356. <https://doi.org/10.1108/LHT-08-2021-0261>
- Mohamed, S., Sethom, K., Namoun, A., Tufail, A., Kim, K.-H., y Almoamari, H. (2022). Customer Profiling Using Internet of Things Based Recommendations. *Sustainability*, 14(18), 11200. <https://doi.org/10.3390/su141811200>
- Mori, K., y Haruno, M. (2021). Differential ability of network and natural language information on social media to predict interpersonal and mental health traits. *Journal of Personality*, 89(2), 228-243. <https://doi.org/10.1111/jopy.12578>
- Murnikov, V., y Kask, K. (2021). Recall Accuracy in Children: Age vs. Conceptual Thinking. *Frontiers in Psychology*, 12, 686904. <https://doi.org/10.3389/fpsyg.2021.686904>
- Nixon, B., y Guajardo, N. R. (2022). The Digital Chameleon: Factors Affecting Perceptions of Convergence in Computer-Mediated Communication. *Journal of Language and Social Psychology*. <https://doi.org/10.1177/0261927X221146143>
- Ouyang, F., Wu, M., Zhang, L., Xu, W., Zheng, L., y Cukurova, M. (2023). Making strides towards AI-supported regulation of learning in collaborative knowledge construction. *Computers in Human Behavior*, 142, 107650. <https://doi.org/10.1016/j.chb.2023.107650>
- Pozdeeva, E., Shipunova, O., Popova, N., Evseev, V., Evseeva, L., Romanenko, I., y Mureyko, L. (2021). Assessment of online environment and digital footprint functions in higher education analytics. *Education Sciences*, 11(6). <https://doi.org/10.3390/educsci11060256>
- Rajesh Kumar, E., Rama Rao, K. V. S. N., Nayak, S. R., y Chandra, R. (2020). Suicidal ideation prediction in twitter data using machine learning techniques. *Journal of Interdisciplinary Mathematics*, 23(1), 117-125. <https://doi.org/10.1080/09720502.2020.1721674>
- Sheikh, S., Patel, M. V., Song, Y., Navuluri, R., Zangan, S., y Ahmed, O. (2021). Social Media Growth at Annual Medical Society Meetings: A Comparative Analysis of Diagnostic and Interventional Radiology to Other Medical Specialties. *Current Problems in Diagnostic Radiology*, 50(5), 592-598. <https://doi.org/10.1067/j.cpradiol.2020.06.001>
- Sjöberg, M., Chen, H.-H., Floréen, P., Koskela, M., Kuikkaniemi, K., Lehtiniemi, T., y Peltonen, J. (2017). Digital Me: Controlling and Making Sense of My Digital Footprint. En L. Gamberini, A. Spagnolli, G. Jacucci, B. Blankertz, y J. Freeman (Eds.), *Symbiotic Interaction* (Vol. 9961, pp. 155-167). Springer International Publishing. [https://doi.org/10.1007/978-3-319-57753-1\\_14](https://doi.org/10.1007/978-3-319-57753-1_14)

- Wang, S., Cui, L., Liu, L., Lu, X., y Li, Q. (2020). Personality Traits Prediction Based on Users' Digital Footprints in Social Networks via Attention RNN. *2020 IEEE International Conference on Services Computing (SCC)*, (PP. 54-56). <https://doi.org/10.1109/SCC49832.2020.00015>
- Yu, W., y Chen, J. (2020). Enriching the library subject headings with folksonomy. *The Electronic Library*, 38(2), 297-315.
- <https://doi.org/10.1108/EL-07-2019-0156>
- Zubareva, S., Zubareva, E., y Pazina, L. (2022). Identification of Students' Professional Competence Based on Big Data and Digital Footprints Based on Big Data Analytics and E-proctoring System. *2022 2nd International Conference on Technology Enhanced Learning in Higher Education (TELE)*, 277-280. <https://doi.org/10.1109/TELE55498.2022.9801042>

**Fecha de recepción del artículo:** 01/12/2022

**Fecha de aceptación del artículo:** 28/02/2023

**Fecha de aprobación para maquetación:** 14/03/2023

**Fecha de publicación en OnlineFirst:** 23/03/2023

**Fecha de publicación:** 01/07/2023