

Herramientas de minería de textos para el análisis cualitativo: Una revisión y análisis documental para el contexto educativo del siglo XXI



Joven investigador: Reinel Andrés Pino Ordoñez
Administrador de Empresas Agropecuarias

Docente :Alexandra María Silva Monsalve
Programa: Ingeniería en Informática

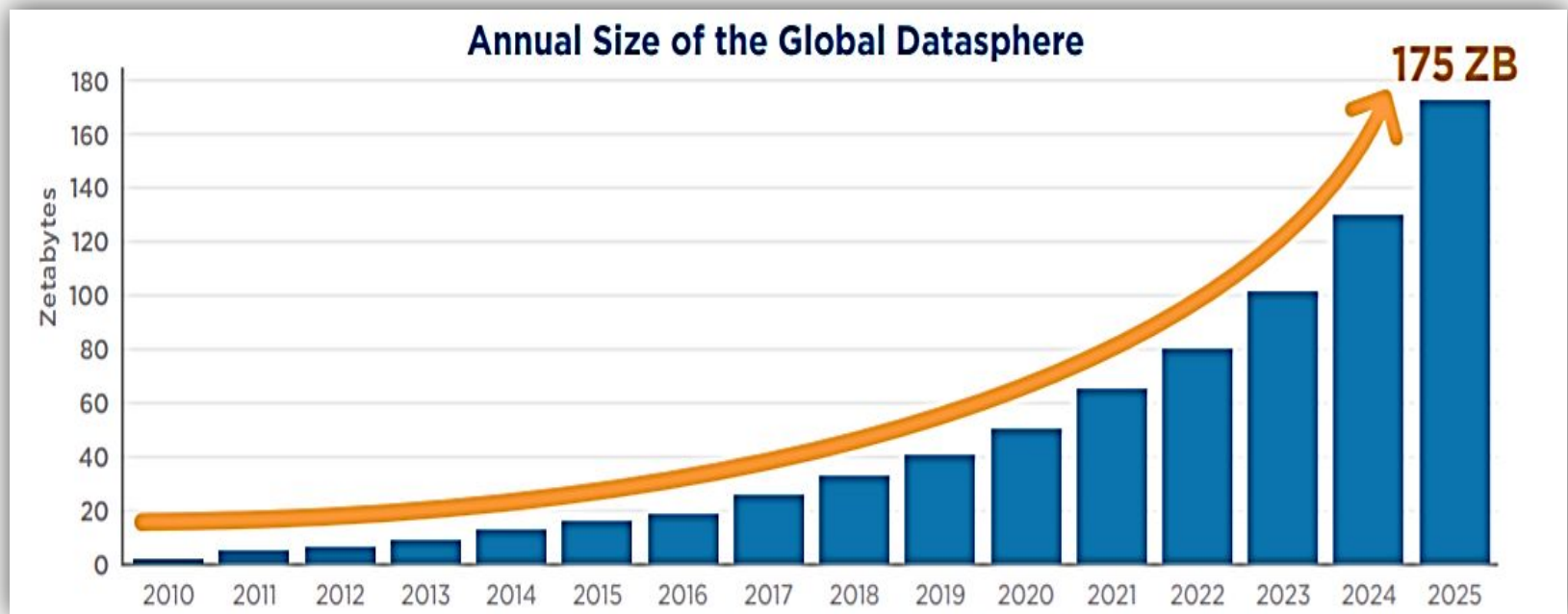
INTRODUCCIÓN

El volumen de información ha crecido de manera exponencial, estos cambios han permitido magnificar y recopilar datos que antes se pensaban eran inaccesibles. Aunado a lo anterior, La Internet, los buscadores, meta buscadores, las bases de datos especializadas y distintos sitios web han generado nuevas fuentes de información que proveen datos más que suficientes para establecer un punto de partida en la revisión de antecedentes o en la construcción de un estado del arte (Internacional Data Coporation IDC, 2018).

OBJETIVO

Implementar herramientas de minería de textos para el análisis de información en recursos y bases de datos electrónicos para la optimización de los estados del arte en trabajos de investigación.





La información se debe orientar hacia dos aspectos:

- (a) Las búsquedas efectivas que permitan la selección de fuentes verificadas y de calidad para abordar las investigaciones (Internacional Data Corporation IDC, 2018).
- (b) La implementación de herramientas que ayuden en la revisión en nivel textual permitiendo la profundización en los contenidos (Internacional Data Corporation IDC, 2018).

Gestión de la Información y el Conocimiento

Sociedad del Conocimiento



La Gestión de la información se define como un proceso investigativo sobre la información, que permite evaluar, presentar, comparar los datos en un determinado contexto (Castells, 1997).

La Gestión del Conocimiento, se orienta desde los procesos y acciones de detección, selección, organización, filtrado, presentación y uso de la información por parte de los sectores de una organización (Torres, 2015). Este último permite transformar los datos, en información y finalmente en Conocimiento (Castells, 1997).



Fases del diseño metodológico en la Investigación Cualitativa

Desde el planteamiento anterior, se identifica la Minería de Textos (MT), en la fase final sobre *Organización y análisis de la información*. Es en esta fase donde se propone utilizar la MT para profundizar en el análisis de contenidos, para apoyar las investigaciones, sobre todo de tipo cualitativo (Hernández, Fernández, & Baptista, 2014).

Minería de Textos (MT)

Es importante diferenciar la minería de textos (*Text mining*) de la recuperación de la información. Esta última consiste en la recuperación automática de fuentes documentales mediante indizaciones de textos, clasificación, categorización. Por otra parte, la minería de textos, se ocupa de un proceso de revisión y análisis en profundidad de fuentes textuales, la cual se apoya desde técnicas como la categorización de texto, el procesamiento de lenguaje natural, la extracción y recuperación de la información o el aprendizaje automático (Rochina, 2017).

Principales técnicas utilizadas en la MT

Pre-procesamiento	Es el proceso mediante el cual los textos se transforman en algún tipo de representación estructurada que facilite su análisis
Representación	La representación depende de la técnica de pre-procesamiento utilizada y determinará el algoritmo de descubrimiento a utilizar
Descubrimiento	Son algoritmos que, a partir de una representación estructurada de la información, son capaces de descubrir regularidades en los textos

Fuente: (Sagrado, 2007)

MÉTODO

El criterio de selección que se tuvo en cuenta, se refiere a la vigencia en tiempo de los documentos y a la pertinencia de los términos de búsqueda (*Mining Text Software*) y *Software* + minería + textos.

#	Bases de datos y redes académicas	Fuentes documentales
1	Dialnet	2
2	ELSEVIER	3
3	Google Académico	7
4	IEEE Explorer	9
5	Portal de revistas Académicas	2
6	Scielo	3
7	Science direct	5
8	Scopus	3
9	Springer	4
10	Taylor & Francis	7

Fuente: Elaboración propia

RESULTADO

S
Herramientas, *software* y servicios para ayuda en el análisis de MT de acceso abierto

#	Nombre	Tipo de Herramienta	Acceso
1	Aika	Página Web/ <i>Software</i>	http://aika.network/resources.html
2	Antconc	<i>Software</i>	https://infotra.wordpress.com/2015/06/25/antconc-software-libre-para-analisis-de-corpus-textuales/
3	GATE	Kit Herramientas <i>Software</i>	https://gate.ac.uk/
4	Textalyser	<i>Software</i>	http://textalyser.net/
5	TextSTAT	<i>Software</i>	https://www.portalprogramas.com/textstat/

Fuente: Elaboración propia

Textalyser

Welcome to the online text analysis tool, the detailed statistics of your text, perfect for translators (quoting), for webmasters (ranking) or for normal users, to know the subject of a text. Now with new features as the analysis of words groups, finding out the keyword density, analyse the prominence of word or expressions. Webmasters can analyse the links on their pages. More instructions are about to be written, please send us your feedback !

Enter your text to analyze here :

or analyze a website :

or select file from local hdd :

Analysis options :

Minimum characters per word :	3
Special word or expression to analyze :	
Number of words to be analyzed :	10
Ignore numbers :	☒
Log the query (only for websites) :	☒
Apply stoplist :	English
Apply own stoplist (separe with blanks) :	
Make a link analysis :	☐
Exhaustive polyword phrases :	☐

Interfaz herramienta en línea *Textalyser*
Fuente: (*Textalyser*,
2019)

Total word count :	162
Number of different words :	122
Complexity factor (Lexical Density) :	75.3%
Readability (Gunning-Fog Index) : (6-easy 20-hard)	22.8
Total number of characters :	1649
Number of characters without spaces :	1380
Average Syllables per Word :	2.48
Sentence count :	9
Average sentence length (words) :	32.25
Max sentence length (words) :	86
(por último la utilización de un dispositivo entrenador i3lap como parte de la solución infotróica se formula desde la mediación tecnológica situada en las teorías constructivistas donde el estudiante puede hacer uso adecuado de la tecnología para generar entornos tecnológicos enriquecidos que permitan a los estudiantes la integración de distintas áreas del conocimiento para adquirir habilidades generales de información y comunicación y nociones científicas para la generación de conocimiento ruíz velasco 2007)	
Min sentence length (words) :	5
(conceptualiza evalúa y aplica)	

Fuente: (*Textalyser*,
2019)

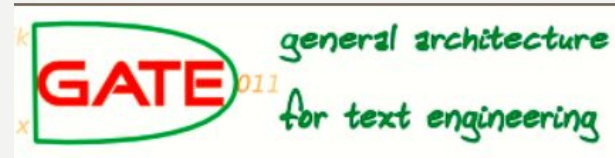
La técnica que utiliza la herramienta se ubica en la frecuencia de aparición de una palabra en un texto, esto permite determinar, la idea principal de un texto.

CONCLUSIONES

Antconc: software libre para análisis de corpus textuales



AntConc



Estilometría



REFERENCIAS

Castells, M. (1997). *The Information Age: Economy, Society and Culture*. Madrid: Alianza Editorial.

Giraldo Zuluaga, W. A. (20 de 03 de 2019). La minería de texto como herramienta para la identificación del perfil investigativo: caso gestión del conocimiento en América Latina de 2012 a 2018. *La minería de texto como herramienta para la identificación del perfil investigativo: caso gestión del conocimiento en América Latina de 2012 a 2018*. Santiago de Cali, Cali, Colombia: Universidad Cooperativa de Colombia. Recuperado el 19 de 05 de 2019, de <http://repository.ucc.edu.co/handle/ucc/7902>

Hammond, K. W., & Laundry, R. J. (2014). Application of a Hybrid Text Mining Approach to the Study of Suicidal Behavior in a Large Population. *IEEE X PLORE*, 8. Recuperado el 31 de 05 de 2019, de <https://ieeexplore-ieee-org.crai-ustadigital.usantotomas.edu.co/document/6758921>

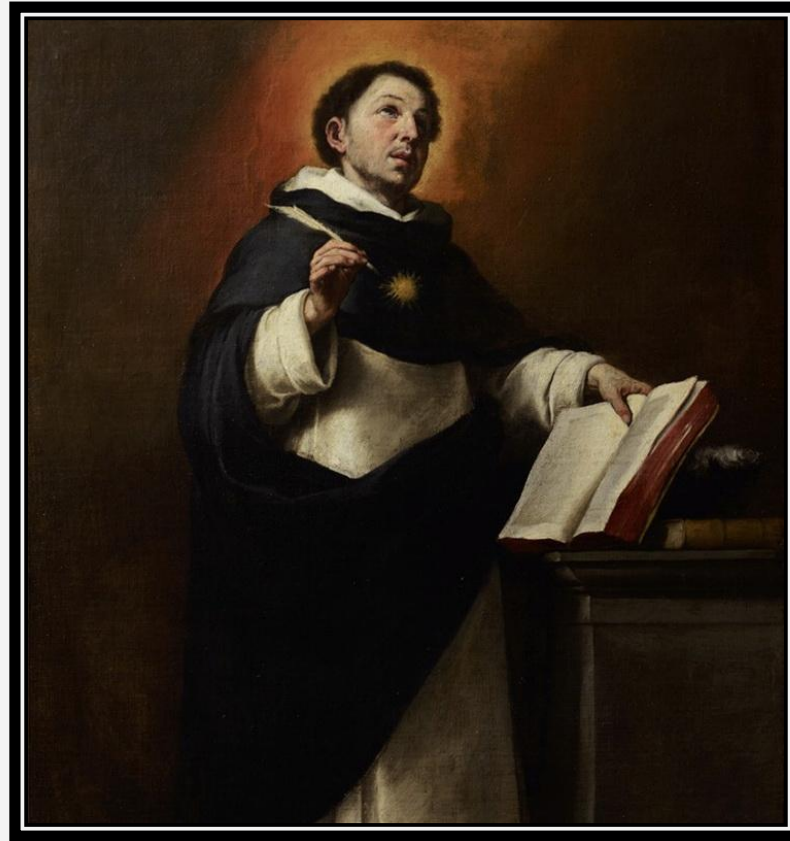
Hernández, R., Fernández, C., & Baptista, L. (2014). *Metodología de la Investigación*. México: Mcgraw-Hill.

Internacional Data Coporation IDC. (2018). *The Digitization of the World From Edge to Core*. Obtenido de <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

Justicia de la Torre, M. (2007). *Universidad de Granada*. Obtenido de [file:///C:/Users/P1ZY/Downloads/233%20\(2\).pdf](file:///C:/Users/P1ZY/Downloads/233%20(2).pdf)

Jia, S. S., & Banggang , W. (2018). Incorporating LDA Based Text Mining Method to Explore New Energy Vehicles in China. *IEEE Xplore*, 6. Recuperado el 29 de 05 de 2019, de <https://ieeexplore-ieee-org.crai-ustadigital.usantotomas.edu.co/document/8502746/authors#authors>

Miller, A. (2018). Text Mining Digital Humanities Projects: Assessing Content Analysis Capabilities of Voyant Tools. *Taylor & Francis*, 20. Recuperado el 26 de 05 de 2019, de <https://www-tandfonline-com.crai-ustadigital.usantotomas.edu.co/doi/full/10.1080/19322909.2018.1479673>



“Lo que se recibe, se recibe al modo del recipiente.”

**GRACIAS POR LA ATENCIÓN
PRESTADA**