



El Profesional de TIC y la Transdisciplinariedad



EDITORES:

- M. en C. Ma. de Lourdes Sánchez Guerrero.
- Dra. Alma Rosa García Gaona.
- Dr. Francisco Javier Álvarez Rodríguez.

EDITORIAL:

ALFA-OMEGA.

ISBN:

978-607-622-368-0

LUGAR:

MEXICO, CIUDAD DE MÉXICO.

FECHA DE PUBLICACIÓN:

DICIEMBRE DE 2017

Índice

PRÓLOGO	5
INTRODUCCIÓN	6
SEMBLANZA DE LOS EDITORES	7
AGRADECIMIENTOS.....	8
COMITÉ REVISOR.....	9
I. PRIMERA SECCIÓN DE ARTÍCULOS DE INVESTIGACIÓN	11
Detección de líneas en imágenes con el algoritmo RANSAC	12
Diseño de un sistema para realizar consultas basadas en contenido en una base de datos multimedia	18
Revisión de técnicas de visualización para Big Data.....	24
Aplicación Móvil para Apoyo Didáctico en la Asignatura de Redes de la DAIS-UJAT	31
Hacia un Sistema de Reconocimiento de Posturas para Personas con Trastorno de Espectro Autista.....	36
Estudio de las Aplicaciones para Reducir el Desperdicio de Alimento	42
Hacia un Sistema de Reconocimiento de Actividades Cotidianas en Humanos para Seguimiento de Actividades Físicas	48
Enriquecimiento de un mapa GIS con otros mapas, haciendo un análisis de posición de atributos geográficos y semántica alineada.	54
II. SEGUNDA SECCIÓN DE ARTÍCULOS DE INVESTIGACIÓN	60
Sistema de Renovación de Préstamos de Libros en Línea.....	61
Experiencia en la implementación de espacios para el aprendizaje de habilidades globales mediada por las TIC como parte de la estrategia de internacionalización de la Universidad de Guadalajara	67
Inteligencia emocional en la enseñanza-aprendizaje de estudiantes de ingeniería de software de la universidad estatal de sonora, UASLRC.	73
Propuesta de una aplicación para Pago Móvil SWPayCard.....	78
Modelado y texturizado de los elementos que integran la recreación virtual de la Universidad Tecnológica de Tabasco	84
Panorama administrativo del uso Moodle utilizando una aplicación web.....	92
Algoritmo para la resolución de ecuaciones con el método de Newton Raphson, implementando Notación Postfija	99
Análisis de factibilidad para la creación de una comunidad virtual orientada a la convivencia de personas sordas y estudiantes universitarios	105
Propuesta de agenda electrónica para control de acceso físico automatizado en espacios académicos de la Facultad de Estadística e Informática, Universidad Veracruzana.....	111
III. TERCERA SECCIÓN DE ARTÍCULOS DE INVESTIGACIÓN	118
Una experiencia institucional con Schoology como Plataforma Virtual Educativa en el Proceso de Enseñanza-Aprendizaje a Nivel Superior.....	119
Oportunidades del Modelo de Aceptación de la Tecnología (TAM) para la aceptación de las Tecnologías de la Información y de las Comunicaciones por parte de adultos mayores.....	128
Sistema de Control de Computadora para Niños con Discapacidad	134
Sistema de Validación y Simulación Sintáctica para la Generación de Intérpretes.....	140
Desarrollo de un MOOC de Pre-cálculo, como un apoyo al programa de Licenciatura en Ciencias Computacionales, en el ICBI-UAEH	146
ProjectHub: Plataforma para Gestión de Proyectos de Software	153
Aplicación para la Extracción de Conocimiento sobre corpus de textos médicos.....	160
Laboratorios Virtuales: migración hacia estándares Web bajo el paradigma de Educación Basada en Web	166
Instrumentación de una red de sensores para el Internet de las cosas	173

Arquitectura de Software para Redes Inalámbricas de Sensores Aplicadas a la Agricultura de Precisión	180
Análisis inteligente del rendimiento académico de redes de estudiantes en un ambiente asistido de tutoría.....	186
Algoritmos Genéticos Aplicados a Problemas de Optimización de Redes de Transporte Público.	193

Revisión de técnicas de visualización para Big Data

Luis Eder Velázquez Peña¹, Lisbeth Rodríguez Mazahua², Giner Alor Hernández³, Beatriz Alejandra Olivares Zepahua⁴ y Silvestre Gustavo Peláez Camarena⁵

Instituto Tecnológico de Orizaba, Avenida Oriente 9 No. 852, Col. Emiliano Zapata, Orizaba, Veracruz, 94320. México

¹levp8915@hotmail.com, ²lrodriguez@itorizaba.edu.mx, ³galor@itorizaba.edu.mx, ⁴bolivares@ito-depi.edu.mx, ⁵sgpelaez@yahoo.com.mx

Resumen. En los últimos 20 años cobró fuerza el término de Big Data, el cual se refiere a los conjuntos de datos que en su tamaño sobrepasan la habilidad de las herramientas de bases de datos típicas para capturar, almacenar, administrar y analizar. El análisis visual de Big Data es un campo nuevo que está emergiendo como una poderosa herramienta para la extracción de información útil. En este artículo se expone la revisión de 83 artículos sobre técnicas de visualización para Big Data de los últimos seis años, para la futura realización de un análisis comparativo de dichas técnicas y determinar cuáles son las más óptimas al momento de hacer un análisis en Big Data.

Abstract. In the last 20 years the term of Big Data took strength, which refers to datasets that in size exceed the ability of typical database tools to capture, store, manage and analyze. Big Data visual analysis is a new field that is emerging as a powerful tool for extracting useful information. This paper discusses the revision of 83 articles on visualization techniques for Big Data of the last six years, for the future realization of a comparative analysis of these techniques and to determine which are the most optimistic when analyzing Big Data.

Keywords: Big Data, técnicas de visualización, visualización Big Data.

1 Introducción

Actualmente el mundo está viviendo la denominada era de la información, donde grandes cantidades de datos se generan y recolectan diariamente tanto para su almacenamiento como para su análisis. Todos estos grandes conjuntos de datos se engloban en el término de Big Data, el cual cobró fuerza en los últimos 20 años. Big Data se refiere a los conjuntos de datos que en su tamaño sobrepasan la habilidad de las herramientas de bases de datos típicas para capturar, almacenar, administrar y analizar [1]. Big Data típicamente incluye masas de datos no estructurados que necesitan analizarse en tiempo real. También proporciona nuevas oportunidades para el descubrimiento de nuevos valores, ayudando a un entendimiento más profundo de los valores ocultos, y además para incursionar en nuevos desafíos [2].

Dado que Big Data requiere ser analizada para extraer información útil, el análisis visual de Big Data surgió como una herramienta para examinar grandes cantidades de datos de forma visual y demostrar una variedad de patrones ocultos, tendencias y correlaciones desconocidas que antes no se percibían o que eran difíciles de observar [3]. Además, al análisis visual toma ventaja de las capacidades de percepción y razonamiento humano para llevar a cabo un análisis exhaustivo de datos, tanto en el panorama general como en niveles detallados [4]. Para realizar el análisis visual de Big Data, se requiere aplicar diferentes técnicas de visualización a los datos. Estas técnicas se usan para la creación de imágenes, diagramas o animaciones para comunicar, entender y mejorar los resultados del análisis Big Data. Presentar la información de tal manera que la gente sea capaz de consumirla de manera efectiva es un desafío fundamental que necesita cumplirse si el análisis de datos es para llevar a una acción concreta [1]. De acuerdo con Han, Kamber y Pei [5], las técnicas de visualización de datos se clasifican en: Orientadas a píxeles, Proyección geométrica, Basadas en iconos, Basadas en jerarquías y grafos, y Relaciones y datos complejos.

Para una mejor comprensión del lector, el presente artículo se divide en las siguientes secciones: la Sección 2 muestra un panorama resumido del estado del arte sobre la problemática planteada, la Sección 3 presenta la metodología utilizada y la Sección 4, los resultados de la revisión de las técnicas de visualización empleadas sobre conjuntos de datos Big Data, por último la Sección 5 plantea las conclusiones y el trabajo futuro sobre esta línea de investigación y, finalmente, se presentan los agradecimientos y referencias.

2 Estado del arte

Al ser un campo nuevo el análisis visual de Big Data, hoy en día no hay muchos reportes técnicos o marcos referenciales [6]–[8] que ayuden a determinar cuáles son las mejores técnicas de visualización para los conjuntos de datos de Big Data. En los análisis existentes no consideran una revisión exhaustiva de la literatura por medio de la que se analicen y clasifiquen los conjuntos de datos y las técnicas de visualización para Big Data.

En [6] los autores plantean una revisión analítica de seis métodos de visualización en Big Data, sin embargo, solo plantean tres características de Big Data (volumen, variedad y velocidad) sin profundizar en otras características (como veracidad y valor), además de no contemplar más métodos o técnicas de visualización.

Lidong Wang et al. [7] también presentan una revisión de la visualización en Big Data, mostrando información similar a la presentada en [6], por lo que no se aporta algo relevante en este campo de investigación.

Finalmente, en [8] se muestra un análisis comparativo de herramientas, tanto comerciales como de código abierto, para la visualización de Big Data, pero no plantean las técnicas de visualización principales o más utilizadas para realizar un análisis de Big Data.

Por tal motivo, este trabajo provee una revisión de 83 artículos de investigación sobre Visualización de Big Data. Esta revisión de la literatura se realizó para proveer una perspectiva de evaluación que permita determinar los campos de aplicación en donde la Visualización de Big Data genera más impacto. Además este trabajo también identifica las técnicas de visualización para Big Data más utilizadas.

3 Metodología utilizada

Para comenzar con el proceso de análisis de las técnicas de visualización empleadas en Big Data, se siguieron los pasos que se muestran en la Figura 1.

Se realizó la búsqueda de artículos en las principales editoriales de investigación, las cuales son ACM, IEEE, Springer y Elsevier, también se tomaron en cuenta otras editoriales siempre y cuando la revista tuviera un factor de impacto.

Una vez ingresado en los repositorios correspondientes, se realizó la búsqueda de los trabajos de acuerdo con los siguientes criterios de búsqueda: palabras clave “*Big Data visualization*” y publicaciones entre 2010 y 2017. Se descartaron artículos de otros idiomas que no fuera inglés y tesis de maestría o doctorado.

Ya teniendo los artículos a analizar, se procedió a buscar los conjuntos de datos utilizados, así como las técnicas de visualización aplicadas a dichos conjuntos de datos. Se agruparon de acuerdo con el dominio de los datos y la clasificación de las técnicas de visualización.

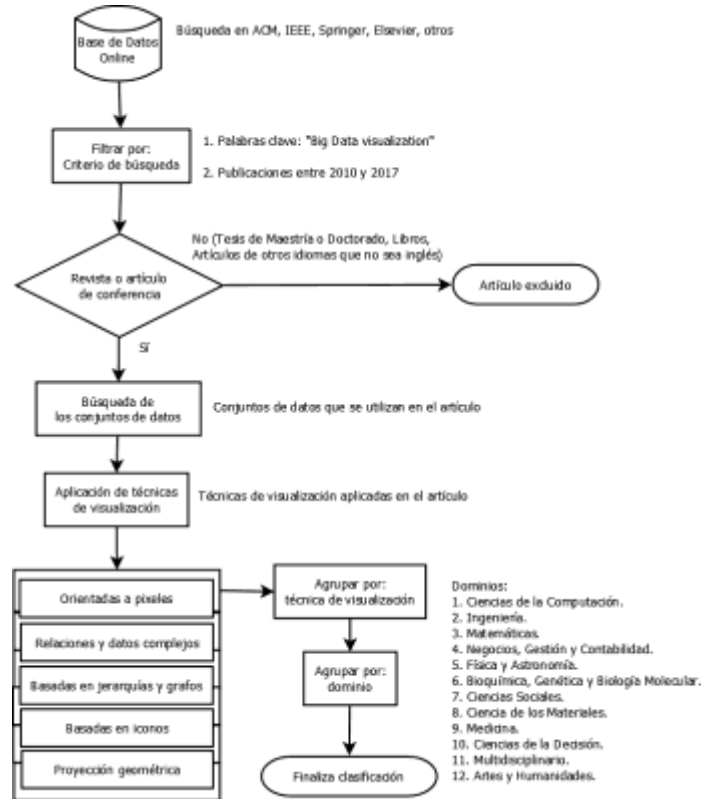


Fig. 1. Metodología de análisis

4 Resultado del análisis

Las siguientes gráficas muestran el resultado del análisis descrito en el punto anterior. Las siguientes clasificaciones ofrecerán importantes tópicos para realizar en un futuro un análisis comparativo de las técnicas de visualización para Big Data.

En la Figura 2 se aprecia cómo aumentó el interés por el análisis visual en el campo de Big Data en los últimos siete años, esto debido al creciente interés por analizar y visualizar datos provenientes de sensores, tanto industriales como geográficos, así como las grandes cantidades de información que se generan en redes sociales e información médica relevante para la prevención de enfermedades. Se contabilizaron un total de 83 artículos relacionados con la visualización de Big Data.

También se distribuyeron los artículos de acuerdo con la editorial de publicación como se aprecia en la Figura 3. De acuerdo con la figura se observa que en la editorial IEEE es donde se concentran la mayoría de los trabajos relacionados con la visualización de Big Data.

Todos los trabajos de investigación de Big Data están enfocados en un dominio específico. Los 108 conjuntos de datos que se encontraron en este análisis de visualización de Big Data, se clasificaron de acuerdo con “*Research Trends*” de Elsevier [9], los cuales son: (1) Ciencias de la Computación, (2) Ingeniería, (3) Matemáticas, (4) Negocios, Gestión y Contabilidad, (5) Física y Astronomía, (6) Bioquímica, Genética y Biología Molecular, (7) Ciencias Sociales, (8) Ciencia de los Materiales, (9) Medicina, (10) Ciencias de la Decisión, (11) Multidisciplinario, (12) Artes y Humanidades. En la Figura 4 se muestra que la mayoría de los conjuntos de datos son referentes a Ciencias de la Computación [10] seguido por el dominio Multidisciplinario [11].

La clasificación de los conjuntos de datos de acuerdo con la técnica de visualización aplicada en sus respectivos artículos se ilustra en la Figura 5. Las técnicas que más se utilizan son las de Relaciones y datos complejos [12] y de Proyección geométrica [13].

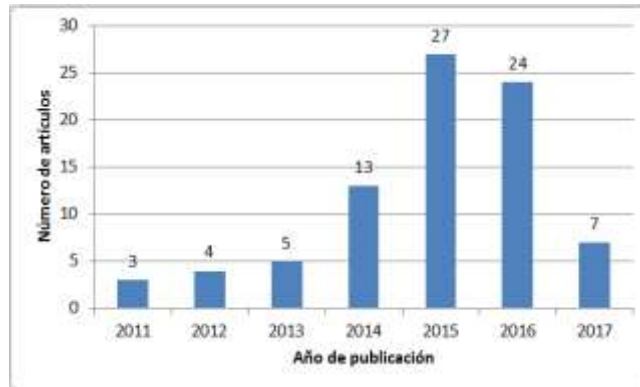


Fig. 2. Distribución de artículos por año de publicación.

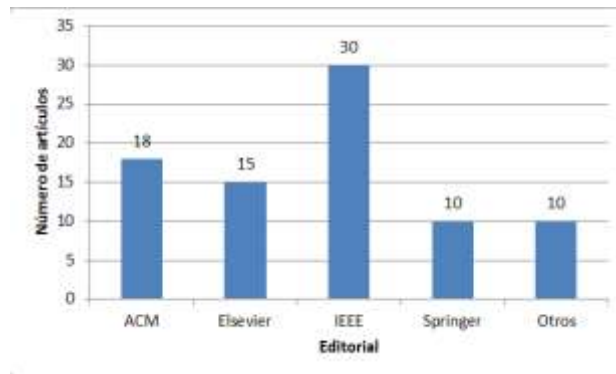


Fig. 3. Distribución de artículos por editorial.

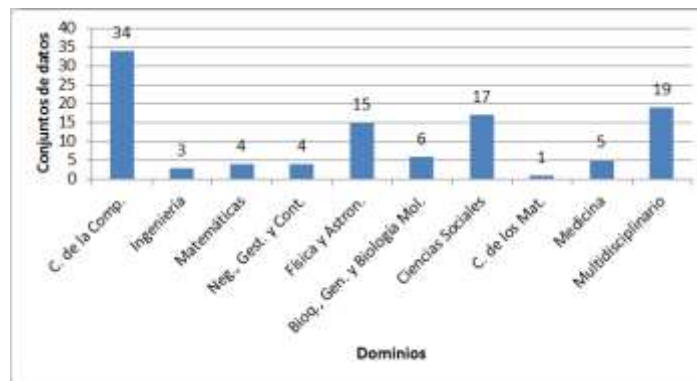


Fig. 4. Clasificación de los conjuntos de datos de acuerdo a su dominio.

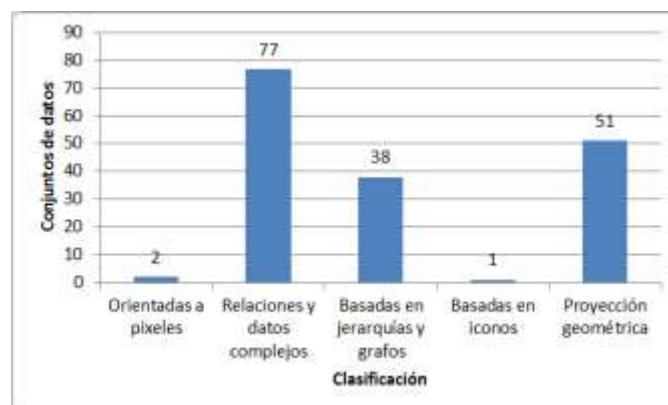


Fig. 5. Clasificación de técnicas de visualización.

La Figura 6 muestra en qué editoriales se concentran los trabajos de acuerdo con la técnica de visualización empleada, donde se observa que la editorial ACM contiene el mayor número de conjuntos de datos en donde

se aplican técnicas de Proyección geométrica. Por otra parte, se aprecia que principalmente en las editoriales ACM e IEEE se concentran los conjuntos de datos en donde se implementan técnicas de Relaciones y datos complejos.

La clasificación de las técnicas de visualización que se aplican a los conjuntos de datos de acuerdo con el dominio de los mismos se muestra en la Figura 7, en donde se observa que para los conjuntos de datos de Ciencias de la Computación se aplican en su mayoría técnicas Basadas en jerarquías y grafos [14], mientras que para conjuntos de Ciencias Sociales y Física y Astronomía se aplican técnicas de Relaciones y datos complejos [15].

Finalmente, se expone la distribución de las herramientas y lenguajes de programación que se encontraron en esta revisión de la literatura. La Figura 8 presenta la cantidad de artículos que usaron cada herramienta y lenguaje, mostrando que la plataforma de Hadoop/MapReduce es la preferida a la hora de realizar análisis de Big Data, dado los beneficios que ésta proporciona al momento de procesar grandes cantidades de datos. Por otra parte, se observa que lenguajes de programación como Python y Java están presentes en el desarrollo de sistemas referentes a Big Data. La sección de “Otros” se refiere a herramientas y lenguajes que sólo se mencionaron en un artículo.

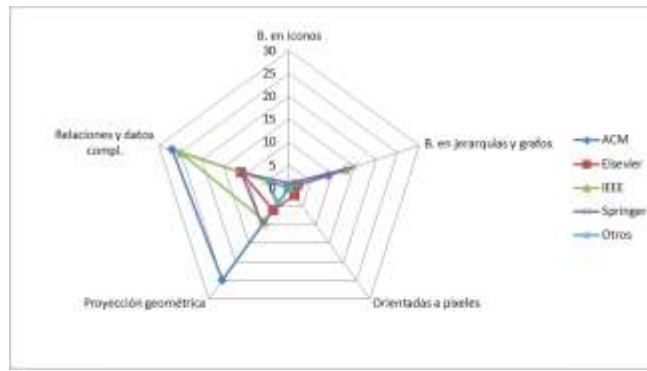


Fig. 6. Distribución de las técnicas de visualización de acuerdo a su editorial.

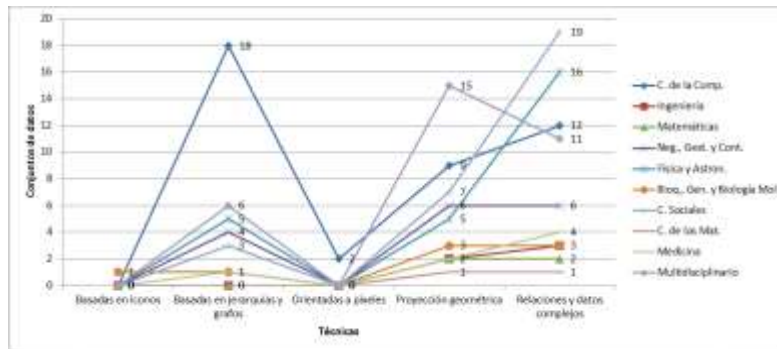


Fig. 7. Clasificación de técnicas de visualización de acuerdo al dominio.

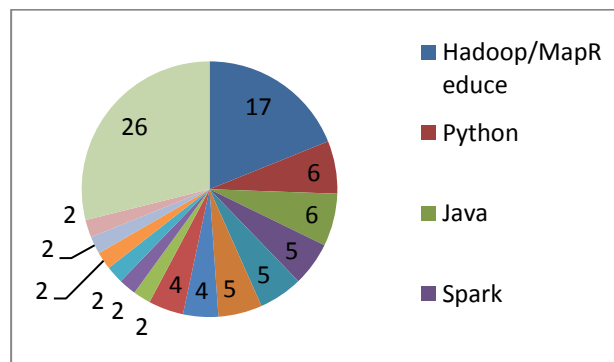


Fig. 8. Distribución de herramientas y lenguajes de programación.

5 Conclusiones y Trabajo a Futuro

Big Data se ha popularizado en los últimos 20 años debido a la creciente explosión en la generación y almacenamiento de los datos, es por esto que requiere analizarse para extraer información útil que le sirva a empresas y a investigadores en la toma de decisiones.

Dado que el análisis visual es un campo nuevo, es importante realizar el análisis comparativo de las diferentes técnicas de visualización para Big Data y así determinar cuáles son las mejores en los diferentes conjuntos de datos masivos, esto mediante un marco de evaluación propuesto que permitirá conocer las ventajas y desventajas de cada técnica de visualización. Dicho análisis pretende beneficiar a una gran comunidad de investigadores Big Data que va en aumento, ayudando en la reducción de tiempo y esfuerzo que cada uno de los investigadores aplica en el análisis visual. Además, la poca literatura que existe sobre este tema no representa una gran ayuda a la hora de determinar qué técnica de visualización aplicar para ciertos conjuntos de datos, por lo que el presente proyecto representa una gran oportunidad para aportar información más completa en este campo.

En este artículo se presentó la metodología y el resultado del análisis de las técnicas de visualización, conjuntos de datos, herramientas y lenguajes de programación que se utilizan en Big Data, mostrando diferentes gráficas que ayudan a entender de una mejor manera la información obtenida y así conocer el estado actual de Big Data respecto a las diferentes áreas de investigación.

Con la información obtenida en la Sección 4 y como trabajo futuro, se procederá a seleccionar los conjuntos de datos más utilizados en Big Data, así como las técnicas de visualización más usadas para realizar un análisis comparativo, aplicando dichas técnicas a los conjuntos de datos y evaluarlas de acuerdo con un marco de evaluación propuesto, y con base en factores como velocidad, precisión, facilidad de interpretación, entre otros, determinar qué técnicas de visualización son mejores para ciertos conjuntos de datos dentro de una análisis de Big Data.

Agradecimientos. Los autores agradecen al Consejo Nacional de Ciencia y Tecnología (CONACYT), al Tecnológico Nacional de México y a la Secretaría de Educación Pública a través del PRODEP por las facilidades prestadas en la elaboración de este proyecto.

Referencias

- [1] James Manyika *et al.*, *Big Data: the Next Frontier for Innovation, Competition and Productivity*. McKinsey & Company, 2011.
- [2] Min Chen, Shiwen Mao, y Yunhao Liu, “Big Data: A Survey”, *Mob. Netw Appl*, vol. 19, pp. 171–209, ene. 2014.
- [3] Aziz Nasridinov y Young-Ho Park, “Visual Analytics for Big Data using R”, presentado en IEEE Third International Conference on Cloud and Green Computing, 2013, pp. 564–565.
- [4] Zeqian Shen, Jishang Wei, Neel Sundaresan, y Kwan-Liu Ma, “Visual Analysis of Massive Web Session Data”, presentado en Symposium on Large Data Analysis and Visualization, Seattle, WA, USA, 2012, pp. 65–72.
- [5] Jiawei Han, Micheline Kamber, y Jian Pei, *Data Mining: Concepts and Techniques*, Third Edition. Morgan Kaufmann, 2012.
- [6] Evgeniy Yur’evich Gorodov y Vasilij Vasil’evich Gubarev, “Analytical Review of Data Visualization Methods in Application to Big Data”, *J. Electr. Comput. Eng.*, vol. 2013, ene. 2013.
- [7] Lidong Wang, Guanghui Wang, y Cheryl Ann Alexander, “Big Data and Visualization: Methods, Challenges and Technology Progress”, *Digit. Technol.*, vol. 1, núm. 1, pp. 33–38, 2015.
- [8] Leishi Zhang *et al.*, “Visual Analytics for the Big Data Era – A Comparative Review of State-of-the-Art Commercial Systems”, presentado en IEEE Conference on Visual Analytics Science & Technology 2012, Seattle, Washington, USA, 2012, pp. 173–182.
- [9] Gali Halevi y Henk Moed, “The Evolution of Big Data as a Research and Scientific Topic: Overview of the Literature”, *Res. Trends*, vol. 30, pp. 3–6, sep. 2012.
- [10] Jian Tang, Jingzhou Liu, Ming Zhang, y Qiaozhu Mei, “Visualizing Large-scale and High-dimensional Data”, presentado en Proceedings of the 25th International Conference on World Wide Web, Montréal, Québec, Canada, 2016, pp. 287–297.
- [11] Ari Wibisono, Wisnu Jatmiko, Hanief Arief Wisesa, Benny Hardjono, y Petrus Mursanto, “Traffic big data prediction and visualization using Fast Incremental Model Trees-Drift Detection (FIMT-DD)”, *Knowl.-Based Syst.*, vol. 93, pp. 33–46, feb. 2016.
- [12] Mingjin Zhang *et al.*, “TerraFly GeoCloud: An Online Spatial Data Analysis and Visualization System”, *ACM Trans. Intell. Syst. Technol. TIST*, vol. 6, núm. 3, p. 24, may 2015.
- [13] Ahmed Eldawy, Mohamed F. Mokbel, y Christopher Jonathan, “HadoopViz: A MapReduce framework for extensible visualization of big spatial data”, presentado en 32nd International Conference on Data Engineering (ICDE), Helsinki, 2016, pp. 601–612.
- [14] Alexandre Perrot, Romain Bourqui, Nicolas Hanusse, Frederic Lalanne, y David Auber, “Large Interactive Visualization of Density Functions on Big Data Infrastructure”, presentado en Symposium on Large Data Analysis and Visualization, Chicago, IL, USA, 2015.

- [15] Wenwu Tang y Wenpeng Feng, "Parallel map projection of vector-based big spatial data: Coupling cloud computing with graphics processing units", *Comput. Environ. Urban Syst.*, vol. 61, núm. Part B, pp. 187–197, ene. 2017.