

Metodología para la aplicación integral de técnicas de visualización en ciencia de datos como apoyo a la toma de decisiones

Methodology for the comprehensive application of visualization techniques in data science as support for decision making

Carlos Román De la Cruz Dorantes^{*}
Ernesto Francisco Bautista Thompson[†]
Jorge A. Ruiz Vanoye[‡]
José Antonio Aguilar Solís[§]

Recibido el 02 de Mayo de 2017. Aceptado el 02 de Agosto de 2017.

Resumen

El avance de la tecnología favorece la generación e incremento de grandes volúmenes de información que rebasan las capacidades cognitiva y de percepción de las personas, creando la necesidad de encontrar nuevas formas de representar dicha información bajo nuevos esquemas que la resuman y presenten de forma más entendible y amigable. La presente investigación busca realizar una valoración de los atributos de ciertas técnicas de visualización con el fin de obtener una jerarquización de la eficiencia de las mismas y al mismo tiempo, pretende identificar problemáticas en la visualización de los datos ofreciendo a los analistas y/o tomadores de decisiones, un marco de referencia que los ayude en la selección de la herramienta de visualización que más se adecúe a sus necesidades.

Palabras Clave: Visualización, Problemas de visualización en la información, Técnicas de visualización.

^{*} Profesor de la Universidad Autónoma del Carmen, Campeche, México y Doctorante de la Universidad Popular Autónoma del Estado de Puebla, carlosroman.delacruz@upaep.edu.mx

[†] Profesor de la Universidad Popular Autónoma del Estado de Puebla, México, eb_thompson@yahoo.com

[‡] Profesor de la Universidad Autónoma del Estado de Hidalgo, México, jorge@ruizvanoye.com

[§] Profesor de la Universidad Autónoma del Estado de Puebla, México, joseantonio.aguilar@upaep.mx

Abstract

The advancement of technology favors the generation and increase of large volumes of information that exceed the cognitive and perceptual capacities of the people, creating the need to find new ways to represent this information under new schemes that summarize and present it in a more understandable and friendly way. The present research seeks to make an assessment of the attributes of certain visualization techniques in order to obtain a hierarchy of the efficiency of them and at the same time, it tries to identify problems in the visualization of the data offering to analysts and / or borrowers of decisions, a frame of reference that helps them in the selection of the visualization tool that best suits their needs.

Keywords: visualization, information visualization problems, visualization techniques.

Introducción

El manejo de grandes volúmenes de información y su representación visual en las empresas se ha vuelto de gran importancia para la toma de decisiones. Existe una creciente necesidad de herramientas de visualización que permitan un mayor aprovechamiento de la información existente en los repositorios de datos, esto a causa de las limitantes que poseen las herramientas actuales empleadas en la visualización de la información.

En la actualidad podemos encontrar una serie de trabajos que soportan la importancia que tiene la visualización de la información hoy en día, este es el caso del trabajo de Gorodov & Gubarev (2013), quienes presentan algunos métodos de visualización para Big Data que dejan ver las ventajas y desventajas de dichos métodos y al mismo tiempo identifican problemas existentes en la visualización de la información, tales como el ruido visual o la pérdida de información. De igual forma, se encuentra la clasificación de la variedad de

métodos de visualización de datos propuesta por Stacey, Salvatore & Jorgensen (2013), donde se busca obtener entre otras cosas una comparación de valores categóricos o entender las relaciones que se dan entre los registros y las gráficas, mismas que son empleadas para trazar y presentar conjuntos de datos con propiedades geospaciales por medio de diversos marcos de mapeo. Existe también una serie de herramientas de software (Pentaho, Flare, ManyEyes, Tableau) con funciones de visualización y con capacidad para manejar gráficos interactivos, pero con limitaciones para generar gráficos provenientes de repositorios de datos que oscilan en el orden de los petabytes o zettabytes (Wang, Wang, G., & Alexander, 2015), esto solo por mencionar algunos. Entonces, existe una necesidad de desarrollo o búsqueda de herramientas de visualización eficientes que permitan la interpretación de resultados de formas más simples o entendibles, sin que exista pérdida de información y se facilite la toma de decisiones (Tascón, 2013). Se dispone de más trabajos realizados que enriquecen la importancia que tiene la visualización como herramienta para la toma de decisiones y donde podemos mencionar los trabajos de otros investigadores como Casner (1991), Bertin (1983) o Nowell y Schulman (2002), quienes opinan que la efectividad de la visualización debe manejarse como una tarea específica.

En Taxonomías de la Visualización de la Información (Guzmán, 2007) se pueden encontrar distintos esquemas de clasificación relacionados con la visualización de la información. Hurtado, Catrillón y Olivar (2013), utilizan la taxonomía para facilitar la identificación de variables asociadas a los datos de interacción de usuarios de software. En Analítica visual en E-Learning (Gómez, García y Therón, 2014), se expone un área de la investigación llamada “analítica visual”, que tiene como objetivo apoyar el razonamiento analítico a través de interfaces visuales interactivas. Bajo el enfoque de evaluación del usuario, las

medidas más comunes de eficacia son el tiempo de finalización de tareas, la tasa de errores, y la satisfacción del usuario. En su estudio Cleveland y McGill (1984) registraron los fallos de la información cuantitativa sobre los gráficos de los sujetos. Por su parte Cox, et al. (1999) y Freedman (2002) se enfocaron en la medición del número de errores en la interpretación de las visualizaciones. Zhu (2007) hace referencia a la necesidad de una metodología que analice sistemáticamente los diversos factores que influyen en la eficiencia y la precisión de la comprensión visual, de tal forma que a través de este análisis, los evaluadores puedan predecir los beneficios de las diferentes características visuales sobre las representaciones no visuales.

A todo lo presentado podemos sumarle las investigaciones realizadas por Rossi & Hirama (2015), Farooq, Mazhar, Khairi & Kamal (2015), Tene & Polonetsky (2012), Fuschi & Tvaronaviciene (2011) y Villards, Olofson & Eastwood (2011), dejando más que claro que construir o buscar herramientas que permitan visualizar datos de manera más eficiente en Big Data, es una tarea compleja.

Hay gran variedad de técnicas de visualización de datos siendo muchas de ellas consideradas como comunes (Line graphs, bar graphs, scatter plots, pie charts, etc.) pero también, existen técnicas visuales poco exploradas o no empleadas que pudieran aportar riqueza al proceso de visualización de la información. Se pretende por medio de este trabajo, desarrollar una metodología bajo la cual se puedan evaluar las capacidades de un conjunto de técnicas visuales dentro del ámbito de la ciencia de los datos, y para esto se requiere de un denso repositorio de datos que permita la aplicación de diferentes técnicas de visualización, su evaluación y comparación, en función de diferentes características y de su eficiencia para visualizar la información asociada a procesos de negocios. Lo anterior

podría dar como resultado, la identificación de una o un conjunto de técnicas de visualización integradas en una metodología para visualización de datos, que permitan mejoras en la representación de los datos y en la identificación de patrones en apoyo al proceso de toma de decisiones.

1. La visualización de datos: características y problemas

Se puede decir que la visualización es la representación y la presentación de los datos que aprovecha nuestras habilidades de percepción visual con el fin de amplificar la cognición (Kirk, 2012). Stacey, Salvatore & Jorgensen (2013), dicen que la visualización de datos consiste en presentar los datos a cualquiera, con el fin de proporcionar una comprensión más intuitiva de los datos o mostrar de una manera para ver una gran cantidad de datos en un área más pequeña. Al final de las cuentas la visualización nos sirve para asimilar información de manera más efectiva mediante la interacción entre nuestros ojos y el cerebro. El éxito de la visualización está asociado con el tipo de técnica empleada en la presentación de los datos. El usuario debe saber exactamente lo que quiere transmitir con la finalidad de que esto lo ayude a seleccionar la técnica más apropiada que pueda usar en la transmisión del conocimiento hacia los usuarios finales. Por lo tanto, hay que tener en mente que las técnicas de visualización tienen características útiles en determinadas áreas y de ahí la importancia de saber elegir la técnica adecuada en base al tipo de proyecto que se esté realizando.

Ahondando un poco sobre el trabajo de Stacey, Salvatore & Jorgensen (2013), en su clasificación propuesta se aprecian las clases de **comparación de valores categóricos** usadas para facilitar las comparaciones entre los tamaños relativos y los tamaños absolutos

de valores categóricos (ej., gráfico de barras), la **evaluación de jerarquías y las relaciones parte-de-un-todo**, las cuales proporcionan un desglose de los valores categóricos en relación a una población de valores o como elementos constituyentes de estructuras jerárquicas (ej., gráfico circular), los **cambios en el tiempo**, empleados para explotar los datos temporales y mostrar las tendencias cambiantes y los patrones de valores durante un período de tiempo continuo (ej., gráfico de líneas, sparklines), el **datos geoespaciales mapping**, que sirve para evaluar las asociaciones, distribuciones y patrones que existen entre los conjuntos de datos multivariantes (ej., el gráfico de dispersión); y finalmente, las **relaciones entre registros y gráficas**, que sirven para trazar y presentar conjuntos de datos con propiedades geoespaciales a través de los diversos marcos de mapeo (ej., mapa choropleth). Esta clasificación nos da una idea de la importancia que tiene saber claramente el tipo de información que estamos manejando, lo que nos ayudará rotundamente a elegir el método más apropiado que nos ayude a transmitir la información de manera más oportuna. Lograr visualizar la información mediante los métodos adecuados ofrece un aporte sumamente importante para las empresas a la hora de tomar decisiones, ya que como dice el dicho, una imagen dice más que mil palabras.

La visualización es la etapa siguiente al análisis de datos y es usada para presentar los resultados obtenidos en ella de manera tal, que el usuario pueda interpretar y tomar decisiones a partir de las gráficas. El tamaño de los datos y su composición juega un papel importante y algunos consejos que se pueden considerar a la hora de seleccionar una herramienta visual podrían ser la necesidad de comprender los datos que está intentando visualizar, tener conciencia del tamaño y la cardinalidad de los mismos, saber con exactitud la información que se quiere transmitir, conocer el público al cual va dirigida la

información y por último, comprender cómo se procesa la información visual; para finalmente lograr transmitir gráficamente la información de la forma más sencilla (Choy, 2012). Entonces, es fácil darse cuenta de los retos presentes en la tarea de visualización dentro del Big Data y se deja ver que a pesar de lo ya logrado, existe una necesidad inherente por explorar nuevas técnicas o desarrollarlas con el propósito de potencializar los resultados que se puedan obtener mediante el análisis de los datos.

Otros aspectos importantes a considerar dentro de la visualización de la información son los problemas a los que actualmente debe uno enfrentarse, y es que con los grandes volúmenes de datos que se generan actualmente, también va en aumento la necesidad de encontrar nuevas formas para visualizar esas cantidades enormes de información. De manera breve podemos mencionar algunos de los problemas encontrados en la visualización de los datos tras una revisión de la bibliografía. Primeramente se puede apreciar que existe una cantidad de métodos tradicionales con limitaciones importantes en sus funciones lo cual les impide explotar apropiadamente los grandes volúmenes de datos; existen dificultades para el análisis y exploración de gráficos de gran tamaño; a lo cual se pueden sumar los problemas de escalabilidad (capacidad de visualización de datos a gran escala), de dinamicidad (evolución o cambio de los datos), el ruido visual, los gráficos imperceptibles y las categorizaciones de las herramientas de visualización de la información que no toman muy en cuenta las necesidades del usuario.

Actualmente la mayoría de los métodos tradicionales de visualización como los histogramas, dot plots, line chart, bar chart, pie chart, area chart, etc. y otros un poco más complejos como treemap o parallel coordinates poseen limitaciones en sus funciones o simplemente no son aplicables bajo ciertas condiciones (no poseen capacidad de manejo de

grandes y variados volúmenes de datos o no ofrecen el manejo de imágenes interactivas).

La escalabilidad y la dinamicidad de los datos dentro de la visualización son para algunos los grandes desafíos a superar en esta disciplina, aunque también se encuentran la diversidad y la heterogeneidad de los datos, como otros problemas de significancia. Existe también una serie de herramientas de software como Pentaho (Hatle, Sayeed, Gupta, Neela & WINTER, 2013), Flare (Flare data visualization, 2016), ManyEyes (Viegas, Wattenberg, Van Ham, Kriss & McKeon, 2007), Tableau (Heer, Mackinlay, Stolte & Agrawala, 2008), con funciones de visualización y con capacidad para manejar gráficos interactivos pero con limitaciones para generar gráficos provenientes de repositorios de datos que oscilan en el orden de los petabytes o zettabytes (Wang, Wang & Alexander, 2015).

En otros estudios como el de Chittaro & Combi (2001), Wiss, & Carr (1998) o el de Daassi, Nigay & Fauvet (2005), se aprecia que las comparaciones empíricas dan poca importancia a la necesidad de un trasfondo teórico y las categorizaciones de las herramientas de visualización de la información no toman muy en cuenta las necesidades del usuario (Wohlfart, Aigner, Bertone & Miksch, 2008). Otros problemas encontrados son la pérdida de valores que genera errores en la lectura de datos, las dificultades de algunas herramientas para reconocer algunos formatos de datos como por ejemplo fechas, y nuevamente aparece el problema del manejo de grandes volúmenes de datos por parte de algunas de las herramientas de visualización, lo cual genera problemas como el bloqueo de la herramienta o la ralentización de la interacción con el usuario. Entre los problemas existentes también se pueden mencionar los relacionados con los gráficos de gran tamaño,

la escalabilidad, la integridad y algunos otros problemas ya citados con anterioridad (Sun, Wu, Liang & Liu, 2013).

Es necesario el desarrollo de estudios que aporten mecanismos de medición de la eficiencia de las herramientas y que faciliten a los analistas la selección adecuada de éstas en base a las necesidades del usuario. Lo anterior deja al descubierto la carencia de herramientas o metodologías integradas que permitan cuantificar aquellas técnicas de visualización que pudieran ser las más adecuadas para un análisis de datos, en base al tipo de datos, proceso de negocio, etc.; no solo en Big Data sino de manera general en el campo de ciencia de datos del cual Big Data forma parte y es un escalamiento en cantidad y diversidad de tipos de datos respecto de la minería de datos. Entonces, son necesarias una o varias metodologías que ayuden al proceso de selección de técnicas de análisis visual de la información.

La complejidad de la tarea de visualización en la ciencia de los datos se debe a la necesidad de generar gráficos que representen información proveniente de repositorios de datos de diversos tipos, repositorios de gran volumen para ciertos casos, la necesidad de nuevas técnicas capaces de lidiar con la variedad de los datos parcialmente estructurados o sin estructura; y por último, la velocidad a la que se genera la información en las organizaciones, lo que genera una demanda de producción de gráficos que permitan ver las relaciones entre las variables involucradas en el análisis de los mismos. Las herramientas visuales actuales poseen debilidades que complican la interpretación de la información, tales como el hecho de que pueden terminar en representaciones imperceptibles, la necesidad de reagrupación de objetos de datos en la pantalla, o el hecho de que los objetos con un peso más pequeño pueden ser opacados por los más grandes generando un diagrama

desastroso; otras herramienta poseen limitaciones en cuanto al número de factores que pueden mostrar a la vez, o se puede dar el caso en donde la visualización dinámica de los datos requiere el cambio de todo el grafico y no solo de la parte afectada. Algunas de ellas no aceptan valores negativos o no son adecuadas para examinar históricos o patrones de tiempo. Son estas algunas de las características que pueden ser valoradas aplicando metodologías de evaluación que apoyen la selección correcta de la herramienta de visualización a la hora de realizar un análisis visual de la información.

2. Propuesta de la metodología para la valoración de las herramientas de visualización.

La investigación considera la búsqueda de técnicas de visualización tanto conocidas como aquellas no utilizadas en áreas relacionadas con la ciencia de los datos, incluida entre ellas el Big Data. Se busca evaluar las técnicas de visualización y su impacto para los procesos de negocios. Es preciso comentar que la revisión de la literatura ha permitido identificar una serie de investigaciones realizadas por diversos autores, las cuales identifican diferentes atributos para caracterizar una técnica de visualización y problemáticas que presentan al representar los datos, y que formarán parte del instrumento que se diseñará como método de valoración para las técnicas empleadas y por emplear. Este proceso de revisión bibliográfica ha permitido también, la identificación de las técnicas de visualización que al momento son usadas para la representación de la información de manera gráfica, pero también ha hecho posible la identificación de otras no empleadas o poco exploradas que podrían aportar beneficios a la tarea de visualización de la información.

Es necesario poder corroborar la riqueza visual que las herramientas seleccionadas pueden aportar en el proceso de toma de decisiones, para lo cual es preciso contar con repositorios de datos representativos de procesos de negocios donde realizar pruebas para validar la funcionalidad de las técnicas. Estos repositorios deberán ser tratados (filtrados, depurados y formateados) con la finalidad de poder valorar en base a las variables identificadas, el alcance de las técnicas en cuestión, y por lo tanto, es necesario el diseño y desarrollo de scripts en lenguaje R o el uso de alguna herramienta como Tableau que agilice la etapa de procesamiento de los repositorios de datos. Entonces, la propuesta metodológica se

Metodología para la aplicación integral de técnicas de visualización en ciencia de datos como apoyo a la toma de decisiones

compone de tres etapas, la primera de ellas corresponde a los requerimientos y comprende las tareas de identificación de herramientas de infovis, atributos, procesos de negocios y repositorios de datos, la segunda etapa es la de valoración y en ella se encuentran comprendida la fase experimental, la de selección de técnicas y la de generación de la matriz de valoración de las técnicas; y por último se tiene la etapa de jerarquización, donde se pretende realizar una clasificación de las técnicas mediante el empleo de los métodos de Mapas Auto Organizados Jerárquicos (GHSOM), el escalamiento multidimensional (MDS), el análisis de componentes principales (PCA) o el proceso analítico jerárquico (AHP). Lo que se pretende es desarrollar una metodología para visualización de información que permita mejoras en la representación de los datos y en la identificación de patrones en apoyo a la toma de decisiones. En la Figura 1 se presenta un diagrama en el cual aparecen las etapas comprendidas para el desarrollo de la investigación.

Figura 1. Esquema metodológico por implementarse.



Fuente: elaboración propia.

Al momento se han hecho pruebas en el ámbito de las redes de computadoras teniendo como objetivo de este primer ensayo, la generación de gráficos que apoyen al administrador de la red en la toma de decisiones relacionadas con la seguridad. En estos primeros intentos, se emplearon los métodos de Análisis de Componentes Principales (PCA) y Mapas Auto Organizados Jerárquicos (GHSOM), para lo cual fue necesaria la generación de matrices de datos estadísticos mediante programas en lenguaje R. Las matrices de datos requeridas, son matrices de m filas x n columnas, donde cada una de las filas corresponde a una máquina de la red y las columnas a un estadístico distinto. El total de columnas pertenecientes a la matriz es de diez y la última pertenece a un número de mapeo asociado a la IP de una máquina. El número de filas dependerá de las máquinas involucradas en el tráfico de la red durante el monitoreo. En la Fig.2 se presenta un fragmento de una matriz de datos generada para el ensayo.

Figura 2. Fragmento de matriz de datos generada para la aplicar PCA y GHSOM

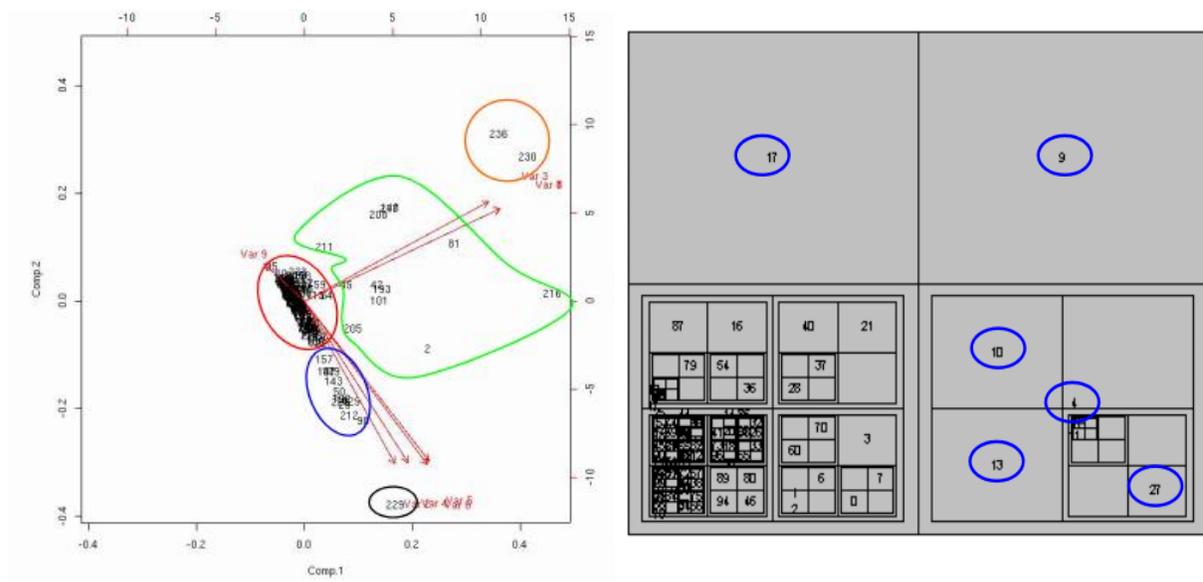
	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	1.977826e-04	0.0000780	3.192797e-07	0.0000059304	0.00000950	0.00008400
[2,]	1.360507e-04	0.0001360	1.581671e-08	0.0000029652	0.00000500	0.00013800
[3,]	8.655136e-05	0.0000850	3.412395e-09	0.0000029652	0.00000500	0.00008700
[4,]	2.384469e-04	0.0002180	8.824531e-08	0.0000088956	0.00003550	0.00025100
[5,]	2.342996e-04	0.0002140	8.197778e-08	0.0000177912	0.00001600	0.00022700
[6,]	1.001875e-04	0.0000960	6.904798e-10	0.0000014826	0.00000300	0.00009800
[7,]	8.531429e-05	0.0000780	9.871603e-09	0.0000029652	0.00000600	0.00007900
[8,]	2.734252e-04	0.0002550	9.097012e-08	0.0000044478	0.00000700	0.00025800
[9,]	1.091042e-04	0.0001120	4.746627e-10	0.0000044478	0.00000500	0.00011500
[10,]	9.114634e-05	0.0000890	9.760280e-10	0.0000059304	0.00000900	0.00009100
[11,]	3.450000e-04	0.0003450	2.048000e-09	0.0000474432	0.00003200	0.00036100

Fuente: elaboración propia.

Una vez generada, filtrada y formateada la matriz de datos, se procedió a la ejecución de los métodos para la generación de las gráficas. En la fig.3 se pueden apreciar los gráficos

(biplots y mapas auto organizados) generados pudiéndose distinguir diferencias entre ellas en la manera de representar la información.

Figura 3. Herramientas gráficas asociadas a la duración de los tiempos de los paquetes de una red de área local.



(a)Análisis Componentes Principales (PCA)

(b)Mapa Auto Organizado Jerárquico (GHSOM)

Fuente: elaboración propia.

La metodología expuesta está alineada para el cumplimiento de los objetivos establecidos en el documento de tesis, los cuales consisten en la creación de un repositorio de datos para la valoración de las técnicas de visualización, la identificación y la evaluación de las fortalezas y debilidades de las técnicas visuales, así como la definición de un nuevo conjunto de herramientas visuales que contribuyan en el análisis de la información y que ayuden al usuario final en la toma de decisiones.

3. Resultados

Del ejemplo presentado se puede concluir que existe un mejor manejo de los datos por parte del método GHSOM al aceptar el uso de valores faltantes dentro de su matriz de datos y aceptar además, valores negativos. De igual forma, la manera de representar la información por parte de GHSOM es más sencilla de entender para el usuario, ya que además de agruparla también la jerarquiza. En la tabla 1, se presenta un sencillo resumen del comparativo realizado entre las técnicas GHSOM y PCA.

Tabla1. Comparativo entre GHSOM y PCA

Características	GHSOM	PCA
Datos incompletos	X	
Valores negativos	X	
Mejor manejo de outliers	X	
Sencillez en sus graficas	X	
Manejo de grandes volúmenes de datos	X	X

Fuente: Elaboración propia

Al momento los resultados obtenidos son derivados de la revisión de la bibliografía y se mencionan de manera puntual a continuación.

- Identificación de variables comunes empleadas en diferentes trabajos de investigación sobre la visualización de la información.
- Identificación de áreas de oportunidad que pueden servir como guía para futuros trabajos.
- Identificación de problemáticas en la visualización de los datos.

- Selección de técnicas para prueba y verificación de su comportamiento en el proceso de representación gráfica de la información.
- Derivado de la revisión bibliográfica, ha sido elaborado y enviado un primer artículo sobre visualización de información.
- Selección de repositorios de datos que cumplan con las características requeridas para la investigación.

De manera general, lo que se pretende mediante este trabajo de investigación es identificar a través de una metodología cuantitativa aquellas técnicas de visualización que de manera individual o combinada, faciliten la tarea de visualización dentro del ámbito de la ciencia de los datos. Estas herramientas deberán ser integradas por medio de un marco metodológico que permita su aplicación en procesos de toma de decisiones y deberán ser capaces de manejar de manera más eficiente, deficiencias tales como la cardinalidad de los datos, la mala representación visual de los datos cuando la correlación entre ellos es fuerte o la velocidad a la que deben generarse. Entonces lo anterior llevará al diseño de una metodología que contenga el conjunto de técnicas, métodos o procedimientos que permitan la aplicación más óptima de las nuevas herramientas de visualización. Una propuesta metodológica que sea capaz de orientar a los futuros usuarios en las tareas de análisis y visualización de la información.

4. Conclusiones

La representación visual de la información es de suma importancia en las empresas para la toma de decisiones. Al momento, se carece de trabajos enfocados a la exploración de las diversas herramientas visuales disponibles que permitan determinar cuál o cuáles de ellas

son capaces de generar aportaciones visuales significativas que puedan ser empleadas por los usuarios finales para la toma de decisiones. También existen trabajos que concuerdan con el hecho, de que hacen falta herramientas de visualización robustas, que posean características básicas para el análisis visual de la información, lo que refuerza la opinión sobre la necesidad de trabajos de valoración de los atributos de las técnicas de visualización, a manera de obtener una jerarquización de la eficiencia de las mismas, tomando en cuenta los problemas vigentes ofreciendo a los analistas y/o tomadores de decisiones un marco de referencia que los ayude a seleccionar la herramienta que más se adecúe a sus necesidades.

Para el ensayo realizado se pudo observar que es el método GHSOM el que posee características más robustas que pueden permitirles a los usuarios una mejor interpretación de la gráfica y por ende, tomar mejores decisiones. Algunas de las particularidades observadas en la herramienta son el manejo óptimo de outliers, el uso de valores negativos, la sencillez en sus gráficos y la capacidad de manejo de datos incompletos. Es preciso realizar experimentos más extensos y complejos que involucren un número más grande de técnicas de visualización con el afán de descubrir las fortalezas y debilidades de cada una de ellas a la hora de generar los gráficos, de esta manera se podría determinar el grado de ayuda que determinada técnica pueden brindar a los usuarios.

Bibliografía

- Gorodov, E. Y. E., y Gubarev, V. V. E. (2013). “Analytical review of data visualization methods in application to big data”. *Journal of Electrical and Computer Engineering*, 2013(22). DOI: 10.1155/2013/969458. Recuperado de: <http://dl.acm.org/citation.cfm?id=2610958>
- Stacey, M., Salvatore, J., & Jorgensen, A. (2013). *Visual Intelligence: Microsoft Tools and Techniques for Visualizing Data*. John Wiley & Sons.
- Wang, L., Wang, G., & Alexander, C. A. (2015). *Big data and visualization: methods, challenges and technology progress*. *Digital Technologies*, 1(1), 33-38.
- Tascón, M. (2013). Introducción: “Big Data. Pasado, presente y future”. *Telos: Cuadernos de comunicación e innovación*, (95), 47-50.
- Casner, S.M. (1991). “A task-analytic approach to the automated design of graphic presentation”. *ACM Transactions on Graphics*, 10(2), 111–151. Recuperado de: <http://dl.acm.org/citation.cfm?id=108361>
- Bertin, J. (1983). *Semiology of Graphics: Diagrams, networks, maps*. Estados Unidos de Norteamérica: University of Wisconsin Press
- Nowell, L., Schulman, R. y Hix, D. (2002). “Graphical Encoding for Information Visualization: An Empirical Study”. *Proceedings of the IEEE Symposium on Information Visualization*, DOI: 10.1109/INFVIS.2002.1173146. Recuperado de: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1173146>

- Guzmán, O. (2007). “Taxonomías de la visualización de la información”. *Sistemas y Telemática*, 5(10), 87-125. Recuperado de: https://www.icesi.edu.co/revistas/index.php/sistemas_telematica/article/viewFile/984/1009
- Hurtado, L., Catrillón, O. y Olivar, G. (2013). “Una Metodología para la Evaluación de Usabilidad de Interfaces de Supervisión Industrial”. *Información Tecnológica*, 24(4), 95-104. Recuperado de: <http://dx.doi.org/10.4067/S0718-07642013000400011>
- Gómez, A., García, F. y Therón, R. (2014). “Analítica Visual en E-Learning”. *El Profesional de la Información*, 23(3), 236-245. Recuperado de: <http://www.elprofesionaldelainformacion.com/contenidos/2014/mayo/03.pdf>
- Cleveland, W. y McGill, R. (1984). “Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods”. *Journal of the American Statistical Association*, 79(387), 531–554. DOI: 10.1080/01621459.1984.10478080. Recuperado de: <http://www.tandfonline.com/doi/abs/10.1080/01621459.1984.10478080>
- Cox, R. (1999). “Representation construction, externalised cognition and individual differences”. *Learning and Instruction*, 9(4), 343–363. DOI: 10.1016/S0959-4752(98)00051-6. Recuperado de: <http://www.sciencedirect.com/science/article/pii/S0959475298000516>
- Zhu, Y., Suo, X. y Owen, G.S. (2007). “Complexity Analysis for Information Visualization Design and Evaluation”. *Advances in Visual Computing*, 4841, Heidelberg, Berlín, Alemania: Springer-Verlag.
- Rossi, R., y Hiram, K. (2015). “Characterizing Big Data Management”. *Issues in Informing Science and Information Technology*, 12, 165-180. Recuperado de: <http://iisit.org/Vol12/IISITv12p165-180Rossi1921.pdf>

- Farooq, M., Waseem, M., Mazhar, S., Khairi, A., y Kamal, T. (2015). "A Review on Internet of Things (IoT)". *International Journal of Computer Applications*, 113(1), 1-7. Recuperado de: [http://www.researchgate.net/profile/Muhammad_Farooq75/publication/273693976_A_Review_on_Internet_of_Things_\(IoT\)/links/5508ac290cf26ff55f83af53.pdf](http://www.researchgate.net/profile/Muhammad_Farooq75/publication/273693976_A_Review_on_Internet_of_Things_(IoT)/links/5508ac290cf26ff55f83af53.pdf)
- Tene, O., y Polonetsky, J. (2013). "Big Data for all: Privacy and user control in the age of analytics". *Northwestern Journal of Technology and Intellectual Property*, 11(5). Recuperado de: <http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1/>
- Fuschi, D. L., y Tvaronavičienė, M. (2011). "Sustainable development, Big Data and supervisory control: service quality in banking sector". *Journal of Security and Sustainability Issues*, 3(3), 5-14. Recuperado de: [http://dx.doi.org/10.9770/jssi.2014.3.3\(1\)](http://dx.doi.org/10.9770/jssi.2014.3.3(1))
- Villars, R.L., Olofson C.W. y Eastwood M. (2011). "White Paper, Big Data: What It Is and Why You Should Care". *IDC Analyze the Future*. 1-14. Recuperado: http://www.admin-magazine.com/HPC/content/download/5604/49345/file/IDC_Big%20Data_whitepaper_final.pdf
- Kirk, A. (2012). "Data Visualization: a successful design process. Packt Publishing Ltd. Choy, J. Visualizing Data Techniques, Including Auto-Charting and Big Data for SAS". *Global Forum 2012*. SAS Institute Inc., Cary, NC, USA.
- Hatle, D. G. N. G. S., Sayeed, A., Gupta, D., Neela, G. P., & Winter, S. (2013). "Pentaho data integration tool". *Business Intelligence Tool*, 2-18.
- Flare data visualization for the web, recuperado en Diciembre 18, 2016, de: <http://flare.prefuse.org/>

- Viegas, F. B., Wattenberg, M., Van Ham, F., Kriss, J., & McKeon, M. (2007). "Manyeyes: a site for visualization at internet scale". *IEEE transactions on visualization and computer graphics*, 13(6), 1121-1128.
- Heer, J., Mackinlay, J., Stolte, C., & Agrawala, M. (2008). "Graphical histories for visualization: Supporting analysis, communication, and evaluation". *IEEE transactions on visualization and computer graphics*, 14(6), 1189-1196.
- Chittaro, L., & Combi, C. (2001). "Representation of temporal intervals and relations: Information visualization aspects and their evaluation". In *Temporal Representation and Reasoning*, 2001. TIME 2001. Proceedings. Eighth International Symposium on (pp. 13-20). IEEE.
- Wiss, U., & Carr, D. (1998). *A cognitive classification framework for 3-dimensional information visualization*. Luleå tekniska universitet.
- Daassi, C., Nigay, L., & Fauvet, M. C. (2005). "A taxonomy of temporal data visualization techniques". *Information-Interaction-Intelligence*, 5(2), 41-63.
- Wohlfart, E., Aigner, W., Bertone, A., & Miksch, S. (2008, July). "Comparing information visualization tools focusing on the temporal dimensions". In *2008 12th International Conference Information Visualisation* (pp. 69-74). IEEE.
- Sun, G. D., Wu, Y. C., Liang, R. H., & Liu, S. X. (2013). "A survey of visual analytics techniques and applications: State-of-the-art research and future challenges". *Journal of Computer Science and Technology*, 28(5), 852-867.
- Freedman, E. y Shah, P. (2002). "Toward a Model of Knowledge-Based Graph Comprehension". *Diagrammatic Representation and Inference*. DOI: 10.1007/3-540-46037-3_3. Recuperado de: http://link.springer.com/chapter/10.1007/3-540-46037-3_3.