

Coordenadas paralelas con filtrado temporal

Andrés Soto Marín y Javier Soto Ledezma

Escuela de Ingeniería,
Universidad Latinoamericana de Ciencia y Tecnología,
ULACIT, Urbanización Tournón, 10235-1000
San José, Costa Rica
[asotom400, jsoto1045]@ulacit.ac.cr
<http://www.ulacit.ac.cr>

Abstract. La visualización básicamente es la representación de datos o información mediante el uso de tablas, imágenes o gráficos como es el caso de nuestra propuesta, cabe resaltar que no solo es la representación también su interpretación de las diferentes tendencias y patrones que se nos hacen invisibles si esos datos o información son presentados en una forma numérica, gracias a la visualización nos permite interactuar, comparar, medir y simplificar la extracción del significado. Coordenadas paralelas son una de las técnicas de visualización más conocidas en el campo de la visualización de la información soportan grandes conjuntos de datos y están hechas para ser creadas o implementadas en sistemas informáticos ya que tienen la característica del brushing, esta característica permite que el usuario final interactúe y manipule el gráfico con el fin de obtener conocimiento, cada columna es una variable así que puede representar muchas variables en un mismo gráfico ellas funcionan de la siguiente forma imaginémonos una tabla cada encabezado de cada columna de la tabla representa el encabezado de cada eje “Y” y cada fila de la tabla representa un trazo dentro del gráfico. En este Proyecto se desarrolló una herramienta que por medio de la visualización y su enfoque en las coordenadas paralelas con filtrado temporal, se realizó una prueba de concepto utilizando los tipos de cambio del Euro con respecto a 41 divisas de diferentes países. El conjunto de datos abarca desde el 04 de enero de 1999 hasta el 05 de noviembre del 2015, la herramienta permite interactuar, comparar y filtrar de acuerdo algún tiempo en específico.

Keywords: Coordenadas paralelas, visualización de la información, Big Data

1 Introducción

En la actualidad y conforme avanzan los días los grandes volúmenes de información y la generación continúa de nuevos datos va en aumento a una velocidad sumamente alta, es por ello que esta investigación se da a la tarea de buscar posibles respuestas a muchas interrogantes. Con este proyecto en el cual el tema principal es la visualización, se pretende facilitar la comprensión de grandes conjuntos de datos temporales y multivariantes, con el fin de apoyar la toma de decisiones en diferentes organizaciones y escenarios.

Por consiguiente creció el interés en diseñar e implementar una prueba de concepto por medio del uso de visualización de información e interacción para facilitar el conocimiento de grandes cantidades de datos temporales y multivariantes. Cabe destacar con respecto a lo investigado, que la obtención de conocimiento de grandes cantidades de datos es compleja, de igual manera los factores temporales en que los datos han sido producidos es un reto difícil de abordar.

Es por esto que el uso de la visualización de la información, en conjunto con las técnicas de interacción, facilita el conocimiento de los datos y contribuye con el proceso de descubrimiento por parte del usuario. Sin embargo, la representación de los factores temporales asociados con el día y hora en que los datos han sido producidos es todo un reto. Por lo que se requiere tener en cuenta que el tiempo agrega una dimensión adicional a la forma de cómo se deben representar los datos de forma visual.

El objetivo general para este proyecto es: Representar y comparar conjuntos de datos temporales y multivariantes mediante el uso de técnicas de visualización e interacción.

Como pregunta de investigación se tiene: ¿Cómo representar y comparar grandes cantidades de datos temporales y multivariantes utilizando visualización e interacción?

Con respecto a la metodología de trabajo se puede citar:

- Investigación.
- Estado del arte.
- Diseño y desarrollo de la herramienta.
- Implementación de una prueba de concepto utilizando los tipos de cambio del Euro con respecto a 41 divisas de diferentes países. El conjunto de datos abarca desde el 04 de enero de 1999 hasta el 05 de noviembre del 2015.
- Prueba de la herramienta mediante escenarios de uso.

2 Estado del arte

La visualización de la información representa y transforma los datos en imágenes para facilitar la extracción de su significado mediante el uso de elementos gráficos, con el fin de explotar la capacidad de percepción visual de las personas para facilitar la extracción de su significado y ampliar el conocimiento (Zhang, 2008).

En la actualidad el volumen de datos que se produce a diario sobrepasa las capacidades de análisis tanto de las personas como de muchos sistemas, por lo que el uso de visualizaciones interactivas y enlazadas permite identificar tendencias y patrones que serían invisibles si son presentados en una tabla numérica. Por otra parte, existe una mezcla de elementos cualitativos, cuantitativos y en algunos casos hasta narrativos, por lo que el uso de visualizaciones permite representar la información de forma simplificada; además de medir, comparar, explorar, descubrir conocimiento e incluso explicar los datos como historias. El objetivo es transformar los datos en información, la información en conocimiento y el conocimiento en inteligencia del negocio para facilitar la toma de decisiones.

En los textos de visualización de información es común que los autores expresen que una imagen vale más que mil palabras (Alcalde, 2015). En ocasiones una palabra, dicho o frase emite sentimientos profundos, pero una imagen puede ayudar a transmitir una mayor cantidad de información, que incluso, un párrafo completo. Otro ejemplo es que realizar una operación matemática compleja de forma escrita puede simplificar su resolución, porque los elementos visuales refuerzan la cognición, mientras que resolverla mentalmente puede ser difícil por no contar con el refuerzo de las imágenes.

La visualización de la información es una especialidad o disciplina que está tomando protagonismo para la representación de grandes cantidades de datos para facilitar su interpretación. La visualización de información hace unos años era utilizada principalmente en la ciencia y era conocida como visualización científica.

Las coordenadas paralelas son usadas para representar grandes conjuntos de datos y es una técnica versátil que permite al usuario final obtener conocimiento (Inselberg, 2009). Este tipo de visualización permite representar información multivariante, con el uso de múltiples ejes paralelos para representar las variables y líneas horizontales para cada elemento en el conjunto de datos: imagine una tabla con columnas y filas, en donde a cada columna se le asigna un eje vertical y a cada fila una línea horizontal, con el valor de cada celda es la escala de representación en el eje vertical.

Si las coordenadas paralelas poseen mucha información es difícil comprender los resultados, por lo que se recomienda el uso de *Brushing* (Kosara, 2006), una técnica de filtrado que facilita la comprensión de la información a simple vista al resaltar varios elementos sobre los demás. Esto facilita la identificación patrones y relaciones que puedan existir en un gran conjunto de datos (Kosara, 2006).

3 Diseño de la visualización

El componente de visualización que se propone en este trabajo se puede ejecutar de forma local accediendo la URL <http://localhost:8000/all>. Dicho componente fue desarrollado en JavaScript y HTML, con MongoDB como motor de base de datos y el uso de los siguientes frameworks o bibliotecas:

1. Express
2. Express-handlebars
3. JQuery
4. JQDataRangerSlider
5. Taucharts

3.1 Detalles del diseño

La Figura 1 muestra una vista general del diseño de la visualización. En el centro de la Figura 1 se pueden observar las coordenadas paralelas, en el lado derecho se muestran los colores que son usados para representar cada elemento, mientras

que en la parte inferior se encuentran un time slider y dos cajas de textos para seleccionar un rango de filtrado. El time slider permite escoger el rango inicial y final del filtrado, lo cual también se puede hacer haciendo clic sobre las cajas de texto de las fechas.

El filtrado de los elementos que aparecen en las coordenadas paralelas se puede realizar haciendo clic sobre los círculos de color en el lado derecho o utilizando *brushing* (ver Figura 2).

4 Escenarios de uso

La interacción de los usuarios con el sistema comienza en la vista general de la visualización, con las líneas de colores representando a cada divisa e indicando el comportamiento de las monedas con respecto al euro. A partir de ahí el usuario selecciona diferentes periodos de tiempo. Conforme con lo anterior, los siguientes dos escenarios son ejemplos en los cuales la visualización de coordenadas paralelas podría ser aplicada.

4.1 Escenario 1

El Banco Central Europeo requiere conocer los tipos de cambio del euro del dólar estadounidense (USD), el real brasileño (BRL), el dólar dólar canadiense (CAD) y el peso mexicano (MXN) para los años 2013 y 2014 con el fin de determinar cuál es el país de América con el tipo de cambio de euro más estable. De acuerdo con esto la Figura 4.1 demuestra que México es el país en el cual el tipo de cambio del euro es el más alto en comparación con los otros países analizados.

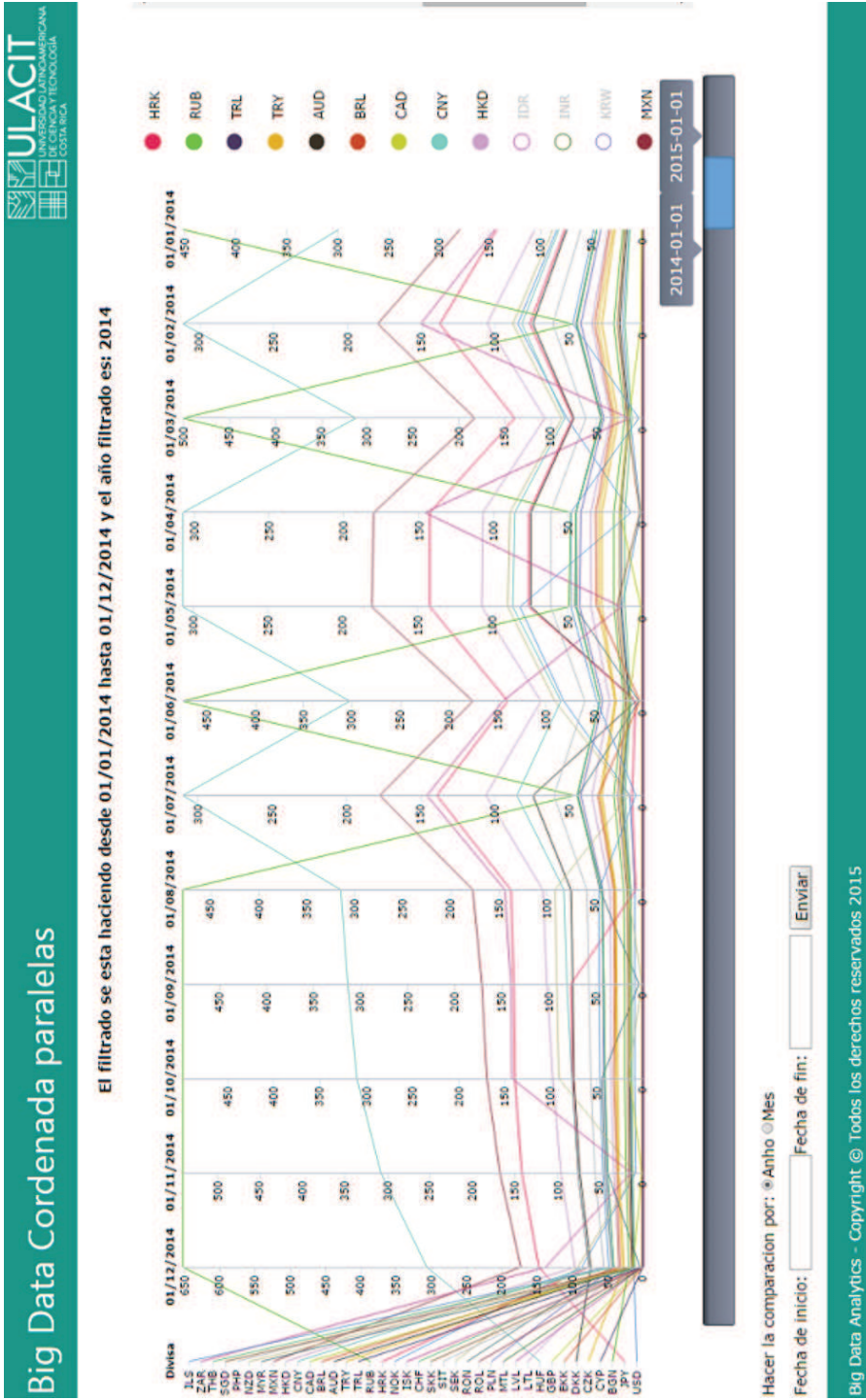


Fig. 1. Vista principal de la visualización gráfico



Fig. 2. Brushing

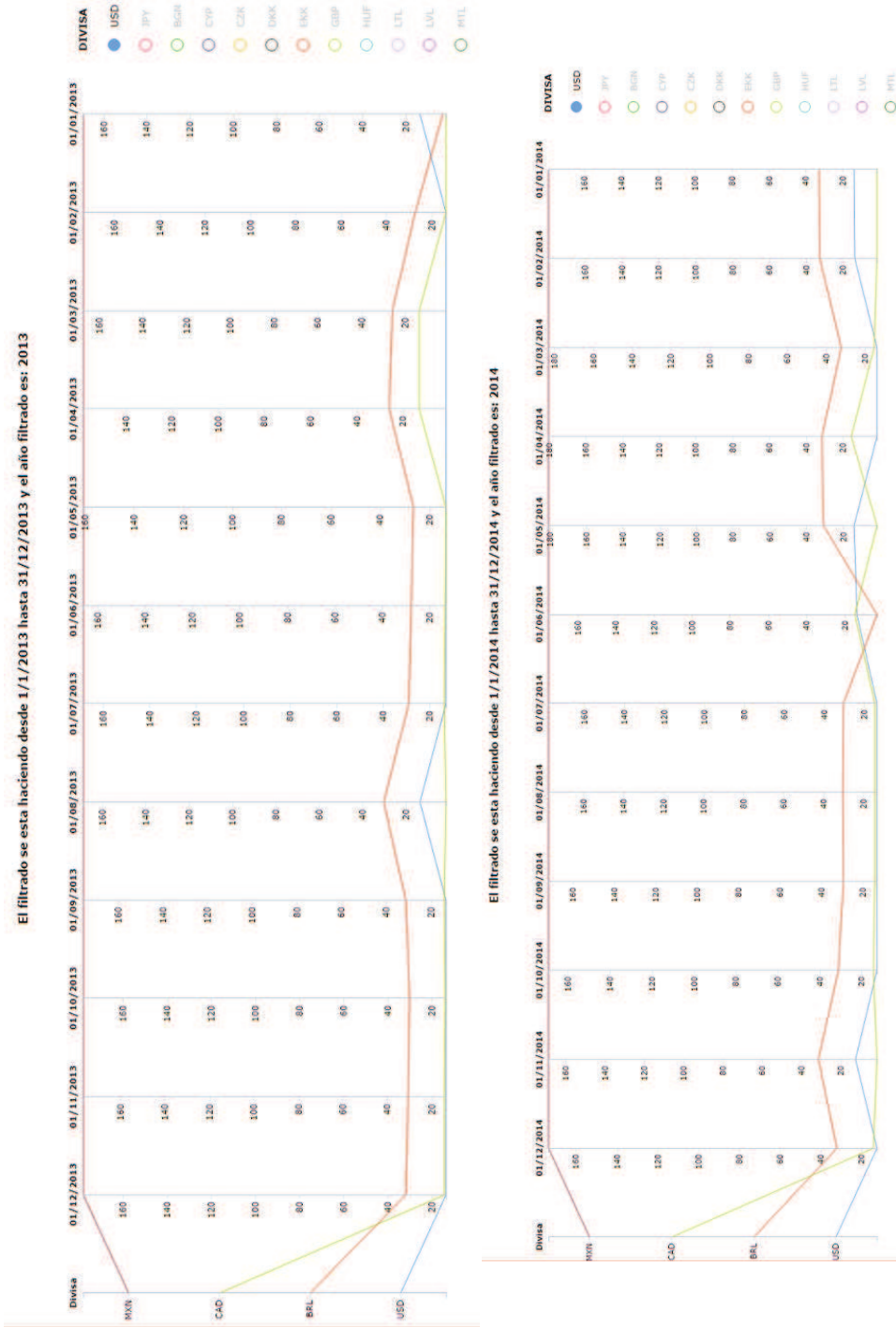


Fig. 3. Tipos de cambios para el dólar estadounidense, dólar canadiense, real brasileño y peso mexicano.

Escenario 2:

Un economista desea conocer el tipo de cambio de euro del franco suizo (CHF), corona noruega (NOK), kuna croata (HRK) y el rublo ruso (RUB) para los meses de enero y febrero de 2015. El objetivo es hacer un pequeño estudio para la compañía en donde trabaja con el fin de demostrar la variación que tiene el euro en los países europeos.

Por medio de la interacción con la visualización el economista puede llegar a diversas conclusiones:

1. El rublo ruso es la divisa más costosa en el periodo en relación con las otras divisas.
2. El franco suizo es la divisa que tuvo el menor costo con respecto al tipo de cambio del euro durante el periodo.
3. La corona noruega y la kuna croata son las divisas que tuvieron más variaciones durante los meses de enero y febrero de 2015.

5 Conclusiones

La meta de este trabajo fue brindar la descripción del proceso y definición de la visualización y su utilización con grandes conjuntos de datos. Sin embargo, no existe una guía detallada que explique de forma precisa cómo llevar a cabo dicho proceso, por el gran número de factores que lo componen.

Debido al alcance limitado del trabajo realizado, existen importantes consideraciones que requieren de más tiempo para poder ser estudiadas a profundidad en un trabajo de esta naturaleza.

En cuanto al diseño propuesto, este permite la comparación de grandes conjuntos de datos de forma intuitiva y rápida, por consiguiente las técnicas de interacción permiten filtrar y obtener detalles adicionales, como por ejemplo, el uso de *brushing*.

En síntesis, la visualización permite obtener conocimiento en poco tiempo haciendo uso de representaciones visuales, interacción y las capacidades cognitivas de los usuarios.

6 Limitaciones

- El diseño actual no permite hacer filtrados con base en los valores de las variables.
- El número de variables, representado por las columnas, se limita al tamaño de la pantalla y la resolución a la cual se encuentra configurada.
- Las tecnologías utilizadas no eran conocidos por el equipo de trabajo de forma previa, por lo que la curva de aprendizaje afectó el desarrollo de características adicionales.

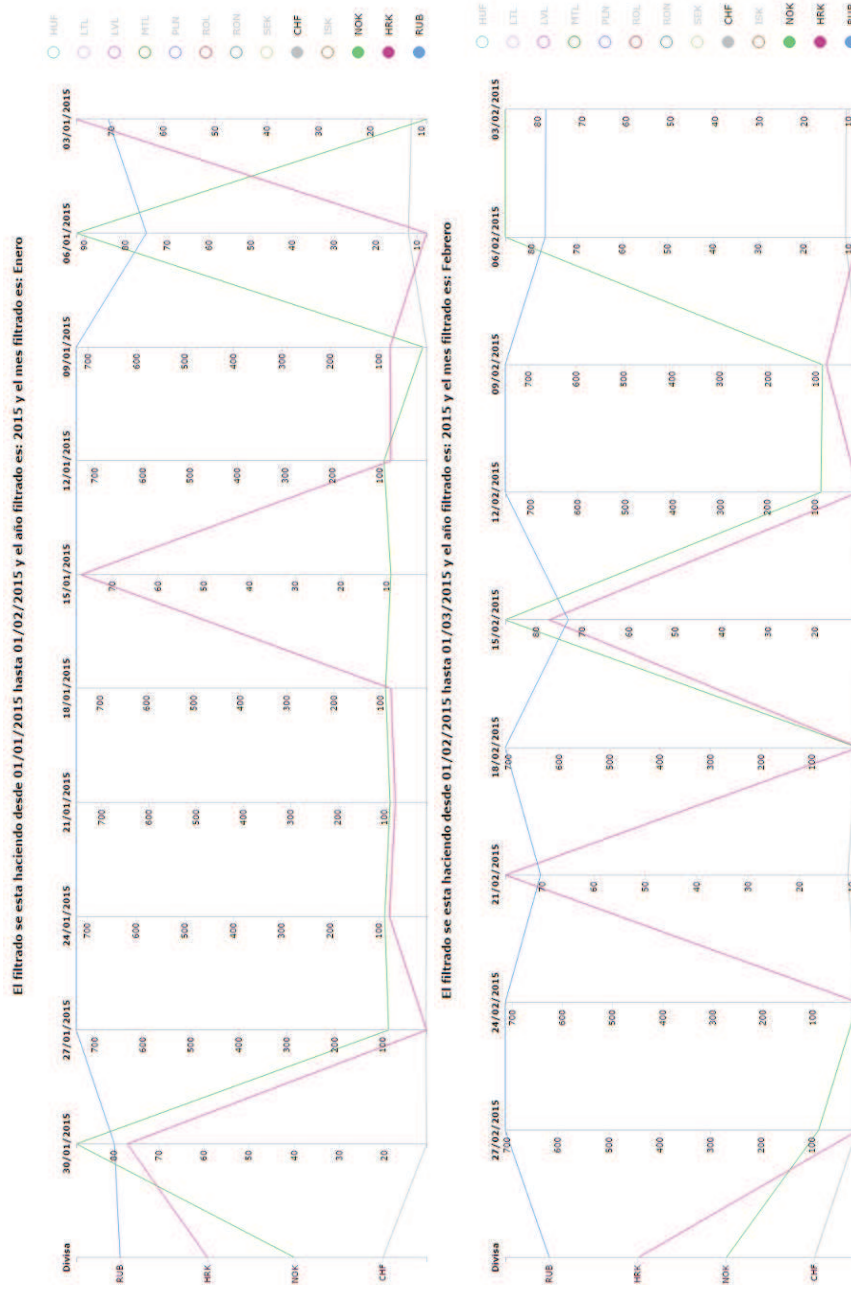


Fig. 4. Variaciones en los tipos de cambio del franco suizo (CHF), corona noruega (NOK), kuna croata (HRK) y el rublo ruso (RUB) para los meses de enero y febrero del 2015.

References

- Alcalde, I. (2015). *Visualización de la información de los datos al conocimiento* (first ed.). Editorial UOC. pages 3
- Inselberg, A. (2009). *Parallel coordinates: Visual multidimensional geometry and its applications*. Secaucus, NJ, USA: Springer-Verlag New York, Inc. pages 3
- Kosara, R. (2006). pages 3
- Zhang, J. (2008). *Visualization for information retrieval*. Springer. (ISBN: 978-3-540-75148-9) pages 2