

Data Viz en Bibliotecas Universitarias

Prog. Luis Olguin, Mag. Alejandra Malberti, Mag. Raúl Klenzi
 Instituto de Informática – Departamento de Informática
 Facultad de Ciencias Exactas, Físicas y Naturales
 Universidad Nacional de San Juan

lolguin@iinfo.unsj.edu.ar; amalberti@gmail.com; rauloscarklenzi@gmail.com

Resumen

Las bibliotecas universitarias generan gran cantidad de datos a partir de sus sistemas de gestión, redes sociales, etc. Sin embargo, la denominada *superabundancia de información* hace que gran parte pase desapercibida y al no tratarla con las herramientas apropiadas, perdamos la oportunidad de encontrar valiosa información oculta en ellos.

En el presente trabajo abordamos la temática de *Data Viz* (Data Visualization) para datos masivos extraídos de una biblioteca universitaria. Entendiendo la visualización como la representación gráfica de información con el doble objetivo: de dar sentido a las decisiones de la biblioteca y ser una herramienta de comunicación con los usuarios finales.

Para ello proponemos el uso de algunas herramientas informáticas para trabajar los datos generados por una biblioteca universitaria y obtener *nuevas vistas gráficas* de la información en ellos contenida

Palabras clave: Data Science, Data Viz, Visualization, Data Analysis, Biblioteca Universitaria.

Contexto

Este trabajo se enmarca en las actividades establecidas en el proyecto “**La Ciencia de Datos en grandes colecciones de datos**” (CICITCA E/1014) **ejecutado** por el Instituto de Informática y el Departamento de Informática de la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan, sujeto a evaluación externa y contenido en el marco del Laboratorio de sistemas inteligentes para la búsqueda de conocimiento en datos masivos.

Introducción

“Existe una cantidad inimaginable de información y datos que constantemente se recogen y se procesan con mayor velocidad. Estamos en el tiempo de los datos masivos pero cuando éstos son demasiados o muy variados hasta el propio analista experto puede perderse en esa lectura, y es en ese momento cuando se hace imprescindible que la visualización de datos sea capaz de mostrar un relato que permita comprender y/o interactuar con los usuarios, clientes o público en general” (Analyze, 2015).

En el trabajo de Olguin, Malberti, Klenzi (2016) se aborda un nuevo enfoque al tratamiento de los datos obtenidos del área de circulación para *modelarlos* con herramientas del área de las *redes sociales* y obtener novedosas *nuevas vistas de los datos transaccionales de una biblioteca universitaria*.

Finch, Jannette (2016) expresa que “la recopilación y visualización de datos en representaciones visuales ayuda a informar al cerebro de forma más rápida y efectiva que la lectura de líneas textuales de información”. Tufte, Edward (2001) enuncia conceptos similares en referencia a la importancia de la visualización, asegura que *los gráficos revelan datos* indicando que *la manera más eficaz de describir, explorar y resumir un conjunto de números es mirar imágenes de esos números*.

En el ambiente de bibliotecas universitarias, donde se centra nuestro trabajo, es posible obtener *miradas más reveladoras* que las simples gráficas generadas por medio de planillas de cálculo que acostumbran nuestros bibliotecarios a explotar.

“La **visualización de datos** es el proceso de representación de datos, en formato gráfico, de una manera clara y eficaz. Ha surgido como una herramienta poderosa y ampliamente aplicable para analizar e interpretar datos grandes y complejos. Se ha convertido en un medio rápido y fácil de transmitir conceptos en un formato universal” (Sadiku, 2016).

Fry, Ben (2007) propone como proceso necesario para la correcta visualización de datos los siguientes pasos:

- **Adquirir:** Recuperar los datos de algún medio magnético.
- **Analizar:** La cantidad de datos que se pueden recopilar y analizar es inmensa. Es necesario poner los datos que se adquieren en una *estructura*. Esta estructura hará que sea más fácil saber transmitir a otros qué datos se tienen (etiquetas).
- **Filtrar:** Eliminar todos los datos que *no son de interés* para la visualización en curso.
- **Minar:** Colocar los datos en un contexto matemático, aplicando técnicas estadísticas o de minería de datos.

- **Representar:** Elegir el mejor modelo visual gráfico para representar los datos.
- **Refinar:** Mejorar la representación visual aplicando algunas herramientas como hojas de estilo (CSS3), gráficos basados en vectores escalables (SVG), etc.
- **Interactuar:** Agregar métodos para controlar las características a ser visibles en la representación visual.

En Finch, Jannette (2016) y Cox, Brian (2012) se generaron visualizaciones de datos para comparar secciones de libros en la biblioteca vs. los gastos en las respectivas áreas, lo que sirvió de punto de partida para discusiones de presupuestos a aplicar. Con los mismos datos se generaron en este trabajo visualizaciones que permitieron detectar obras para expurgo, relaciones entre obras y usuarios, etc.

El proyecto Seattle Central Library (2014) es una interesante muestra de que la exposición de datos en formatos gráficos no tradicionales y dinámicos le brinda una *nueva vista* a los usuarios de la biblioteca. Mediante el uso de pantallas LED ubicadas en el área de circulación de la biblioteca, se ofrecen vistas gráficas de la evolución de los préstamos, áreas de más interés (Dewey), listas de palabras claves de las obras solicitadas, etc. Ésta abundante información, que se procesa cada 60 minutos, es un atractivo entre los servicios que ofrece esta biblioteca.

Mediante la aplicación de estas *vistas de los datos* se busca que la biblioteca las utilice para la toma de decisiones y como elemento orientativo para los usuarios.

El desafío es determinar cómo se aplicará la visualización, pasando del solo hecho de representar datos de una *manera distinta*, a guiarnos en cómo seleccionar los datos a desplegar para que representen información.

Líneas de Investigación, Desarrollo e Innovación

La posibilidad de almacenar enormes volúmenes de datos exige el estudio de nuevas formas de representarlos para *extraer información valiosa*.

En este trabajo se propone el estudio de herramientas y técnicas que permitan analizar grandes volúmenes de datos y generar representaciones visuales que posibiliten la toma de decisiones tanto a los bibliotecarios como a los usuarios finales de una biblioteca universitaria.

La línea principal de investigación, que se ajusta a una investigación en informática aplicada, consiste en el empleo de soluciones, preponderantemente open source o freeware, para la toma de decisiones en el ámbito de una biblioteca universitaria.

Resultados y Objetivos

Se propone la aplicación de herramientas informáticas que permitan el tratamiento de grandes volúmenes de datos y su posterior vista gráfica (*Data Viz*) como otra forma de *descubrir información* oculta en los datos de una biblioteca universitaria.

Se espera que al implementar estas herramientas informáticas en el ámbito de la biblioteca se pueda establecer, mediante el análisis comparativo, cuales son las más apropiadas desde un enfoque de HCI (interacción humano computador), entendiendo que los usuarios finales deben ser la referencia para el diseño de la interfaz de un sistema y este debe adecuarse a la forma de interactuar de ellos.

Formación de Recursos Humanos

En esta instancia se está dirigiendo una Tesis de Maestría en Informática, orientada a la aplicación de herramientas de Ciencia de Datos y Redes Sociales en el contexto de bibliotecas universitarias.

De manera continua se transfieren los conocimientos adquiridos a las cátedras de las carreras del Dpto. de Informática de la FCEF y al equipo de trabajo que forma parte del **Laboratorio de Sistemas Inteligentes para la Búsqueda de Conocimiento en Datos Masivos** (Instituto de Informática, Resol. 02-2015), cuyo objetivo es la extracción de conocimiento en grandes bases de datos mediante la utilización de algoritmos de minería de datos y aprendizaje de máquina soportado por arquitecturas secuenciales y paralelo-distribuidas.

Referencias

Analyze (2015). Data Visualization. *Recuperado de* <http://www.analyze.cl/data-visualization/>

Cox, Brian (2012). *Discovering the impact of library use*. Educause Review Online. *Recuperado de* <http://er.educause.edu/articles/2012/7/discovering-the-impact-of-library-use-and-student-performance>

Finch, Jannette (2016). Using Data Visualization to Examine an Academic Library Collection. *College & Research Libraries*, Noviembre, 165-178.

Fry, Ben (2007). *Visualizing Data: Exploring and Explaining Data with the Processing Environment*. O'Reilly Media, Inc

Gutsche, Betha (2016). *Data Data Everywhere – Finding Focus in a Sea of Visualizations*. *Recuperado de* http://www.wla.org/assets/Alki/Alki_Mar2016_v32-1.pdf

Olguin, Malberti, Klenzi (2016). *Red social de coprestamo en bibliotecas universitarias*. Workshop de Investigadores en Ciencia de la Computación. Entre Ríos.

Sadiku, Matthew (2016). *Data Visualization brief*. International Journal of Engineering Research And Advanced Technology(IJERAT).Volume. 02 Issue.12. p. 11-16.

Seattle Central Library (2014). *Making visible the invisible*. Recuperado de <http://www.mat.ucsb.edu/~g.legrady/gWeb/Projects/spl/spl.html>

Tufte, Edward (2001). *The Visual Display of Quantitative Information* (2^a. ed.). Cheshire, Conn.:Graphics Press.