

Universidad Tecnológica del Centro de Veracruz - Innovasolver

Proyecto: Plataforma tecnológica de georecomendación para productos y servicios en modalidad subasta inversa basado en técnicas de cómputo sensible al contexto, análisis de sentimientos, minería de opiniones.

Entregable: (A4.5) Definición de técnicas de visualización grafica para el análisis de datos almacenados en repositorios Big Data

ÍNDICE

ILUSTRACIONES.....	3
INTRODUCCIÓN.....	5
TÉCNICAS DE VISUALIZACIÓN DE DATOS	7
1. TÉCNICAS ORIENTADAS EN PROYECCIÓN	7
a. PROYECCIÓN LINEAR	7
b. PROYECCIÓN NO LINEAR	9
2. TÉCNICAS BASADAS EN GRAFOS	10
a. GRÁFICOS BÁSICOS (BASIC GRAPHS)	10
b. SPECIFIC GRAPHS	12
3. TÉCNICAS BASADAS EN ICONOS.....	13
a. STICK FIGURE ICON	13
b. SISTEMAS	14
4. TÉCNICAS BASADAS EN JERARQUÍAS	15
a. DIMENSIONAL STACKING	16
b. TREE MAP	16
5. TÉCNICAS BASADAS EN PIXEL.....	18
a. TÉCNICAS INDEPENDIENTES DE CONSULTA (QUERY-INDEPENDENT TECHNIQUES)	18
b. CHERNOFF FACES	19
c. TÉCNICAS DEPENDIENTES DE CONSULTA (QUERY-DEPEND TECHNIQUES)	19
CONCLUSIÓN.....	23
BIBLIOGRAFÍA.....	25

ILUSTRACIONES

Ilustración 1. Clasificación de las técnicas de visualización de datos (Keim, Kriegel, & Ankerst, 1995).	5
Ilustración 2. Clasificación de las técnicas de visualización de datos (Keim D. A., 2000).	6
Ilustración 3. Ilustración del PCA para datos tridimensionales de expresión genética, localizados en un subespacio bidimensional (Scholz, 2006).	7
Ilustración 4. Ejemplo de la aplicación del análisis LDA con datos, espacio, puntos de prueba, y región de decisión (Ganapathiraju & Balakrishnama, 1998).	8
Ilustración 5. Conjunto de datos artificiales: rollo suizo (Swiss roll, izquierda) y caja abierta (open box, derecha) (Hujun, 2003).	9
Ilustración 6. Comparación de diferentes técnicas de proyección no lineal (Nan & Weiwei, 2016).	9
Ilustración 14. Clasificación de las técnicas basados en grafos según Daniel A. Keim (Keim D. A., 2000).	10
Ilustración 15. Dibujos de gráficos básicos o en 2D, según Daniel A. Keim.	11
Ilustración 16. Gráfico ortogonal (izquierda) y gráfico simétricamente optimizado (derecha) (Keim D. A., 2000).	11
Ilustración 17. Técnica de visualización de información HyGraph (Keim & Kriegel, 1996).	12
Ilustración 18. Ejemplo de implementación de un DAG. Tomado de Creative Proteomics.	13
Ilustración 19. Ejemplo de grafos DAG.	13
Ilustración 8. Técnica de visualización "Stick Figure", mostrando un icono único y una familia de iconos (Keim D. A., Visual Techniques for Exploring Databases, 2000).	14
Ilustración 9. Visualización de los datos del censo a través de la técnica Stick Figure. (Keim & Kriegel, Visualization Techniques for Mining Large Databases: A Comparison, 1996)	14
Ilustración 20. Pantalla del software Tom Sawyer.	15
Ilustración 10. Clasificación de las técnicas basadas en jerarquías.	15
Ilustración 11. Técnica "Dimensional Stacking" (Keim D. A., 2000).	16
Ilustración 12. Técnica Treemap, representada en un diagrama Venn (Johnson & Shneiderman, 1993).	17
Ilustración 13. Visualización de 1000 archivos, a través de la técnica treemap (Johnson & Shneiderman, 1993).	17
Ilustración 21. Clasificación de las técnicas basadas en pixel (Pixel-oriented techniques).	18
Ilustración 22. Arreglo completamente recursivo (Keim, Kriegel, & Ankerst, Recursive Pattern: A Technique for Visualizing Very Large Amounts of Data, 1995).	18
Ilustración 7. Técnica Chernoff Faces. (Morris, Ebert, & Rheingans, 2000)	19
Ilustración 23. Proceso de cálculo para visualizaciones (Keim D. A., Visual Techniques for Exploring Databases, 2000).	20

Ilustración 24. Técnica de espiral (Spiral Technique) (Keim D. A., Visual Techniques for Exploring Databases, 2000).	21
Ilustración 25. La técnica de serpiente-espiral. (Keim D. A., Pixel-oriented Visualization Techniques for Exploring Very Large Databases, 1996)	21
Ilustración 26. Posicionamiento de los datos en la técnica de ejes (Keim D. A., Pixel-oriented Visualization Techniques for Exploring Very Large Databases, 1996).	22
Ilustración 27. Herramientas de visualización no empleadas según Ana Cristina Umaquina, et al.	23
Ilustración 28. Técnicas de visualización de Big Data más utilizadas, según Ana Cristina Umaquina, et al (Umaquina, Peluffo, & Alvarado, 2016).	24

INTRODUCCIÓN

En México existen 54,068791 personas económicamente activa en el 2017 según el INEGI. Cada una de las personas genera una gran cantidad de información en diferentes instituciones públicas, privadas o de gobierno, y deben existir diferentes herramientas tecnológicas para el tratamiento de grandes cantidades de información. En este sentido las técnicas de visualización para grandes cantidades de información, permiten analizar y visualizar los datos de manera que sea fácil de interpretar para los interesados.

El término Visual Data Mining se refiere al proceso de búsqueda y análisis en grandes cantidades de información de aquellos datos que son potencialmente usable y de interés, con base en los parámetros. Uno de los pioneros en la exploración de técnicas de visualización de datos fue Edward R. Tufte con la publicación de su libro “Envisioning Information” en 1990 en donde se centra en: 1) la visualización de datos con semántica 2D y 3D, 2) desarrollo de técnicas de visualización para diferentes tipos de datos a través de un modelo físico y 3) desarrollar técnicas de visualización para datos de múltiples dimensiones. (Keim D. A., 2000). A partir de los trabajos anteriores Daniel Keim, realiza la clasificación mostrada en la ilustración 1.

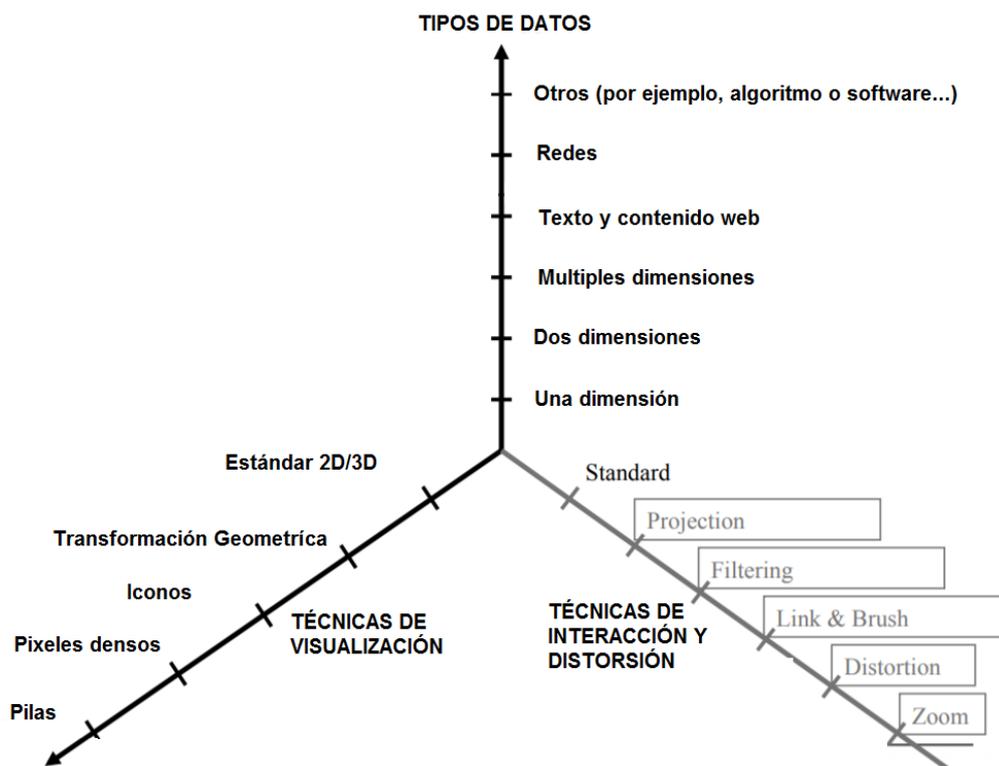


Ilustración 1. Clasificación de las técnicas de visualización de datos (Keim, Kriegel, & Ankerst, 1995).

En el presente documento se abordarán diferentes técnicas de visualización de datos, las cuales brindan un alto valor en la exploración y análisis de datos. Estas técnicas permiten buscar información relevante dentro de grandes cantidades de datos, dado que, a mayor cantidad de tiempo, los datos se van a incrementar de manera potencial, y cada vez resulta más complicado el análisis de los datos. Las técnicas abordadas son las orientadas en proyección, basadas en iconos, basadas en jerarquías, basadas en grafos, y basadas en pixeles, como se muestra en la ilustración 2.

Técnicas de visualización de datos	Basadas en iconos	Chernoff Faces
		Stick Figure Icon
	Basadas en Jerárquías	Dimensional Stacking
		Tree Map
	Basadas en grafos	Básicos
	Específicos	
	Sistemas	
Basadas en pixel	Independiente de consulta	
	Dependientes de consulta	
Orientadas en proyección	Linear	
	No linear	

Ilustración 2. Clasificación de las técnicas de visualización de datos (Keim D. A., 2000).

TÉCNICAS DE VISUALIZACIÓN DE DATOS

El término de Big Data, abarca a todos aquellos conceptos involucrados con la gran cantidad de datos (volumen, variabilidad, y velocidad). En español es conocido como **Macrodatos**, el cual define Mario Tascón como “Todo aquello que tiene que ver con grandes volúmenes de información que se mueven o analizan a alta velocidad y que pueden presentar una compleja variabilidad en cuento a la estructura de su composición” (Tascón, 2013).

Para las empresas, la información se ha convertido en el recurso vital de la misma, y como el crecimiento de la misma es exponencial en un año, es importante para las empresas determinar una opción para almacenar y gestionar dicha información, es aquí donde surge el valor del **Big Data**. Para la implementación de un sistema basado en Big Data considerando el volumen de la información, la variabilidad o cambio de la información y la velocidad en la que se va generando dicha información, se debe considerar que se está trabajando con alguno de los motivos anteriormente mencionados o mejor conocidos como 3V's. Dentro del análisis de grandes volúmenes de información, existen diferentes técnicas que permiten el análisis visual de datos en donde se integran métodos computacionales automáticos para una mejor comprensión y análisis de datos complejos.

1. TÉCNICAS ORIENTADAS EN PROYECCIÓN

Son descritas como aquellas técnicas que permite la visualización de documentos, a través de un procedimiento de reducción de dimensiones. Estas técnicas, son separadas en proyección lineal y proyección no lineal.

a. PROYECCIÓN LINEAR

En la proyección lineal, se contemplan dos técnicas de análisis.

- i. La primera es el análisis de componentes principales (**PCA**, por sus siglas en inglés de **Principal Component Analysis**), la cual consiste en el análisis de variables cuantitativas, dependientes y correlacionadas, las variables contienen las observaciones de una tabla de datos.

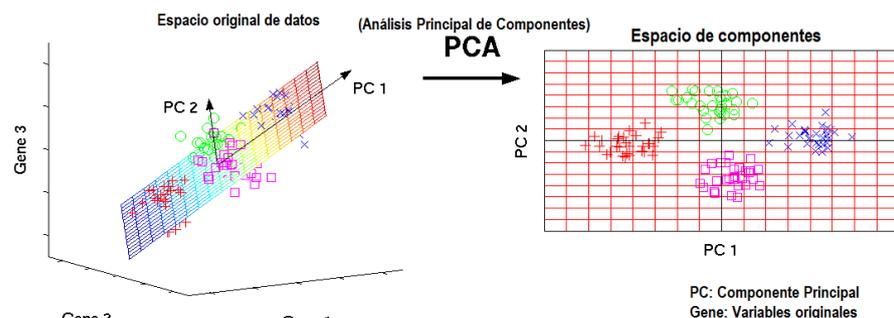


Ilustración 3. Ilustración del PCA para datos tridimensionales de expresión genética, localizados en un subespacio bidimensional (Scholz, 2006).

PCA lo que busca es extraer la información importante de una tabla de datos (data table), en donde dicha información es almacenada en nuevas variables ortogonales, a las que se les conoce como *componentes principales*. PCA usa las variables y las observaciones para representarlas a través de un patrón la similitud entra las mismas, mediante puntos en un mapa. En términos generales, el *PCA es usado para visualizar datos mediante la reducción de la dimensionalidad de los datos* (Scholz, 2006).

- ii. El segundo es el análisis discriminante lineal (**LDA**, por sus siglas en inglés de **Linear Discriminant Analysis**), el cual fue creado por Ronald Fischer en 1936 y consiste en minimizar la varianza que existen dentro de la clase y la dispersión entre las mismas, construyendo una región de decisión definida entre las clases, como se aprecia en la ilustración 4. El LDA es un método de clasificación supervisado de variables cuantitativas.

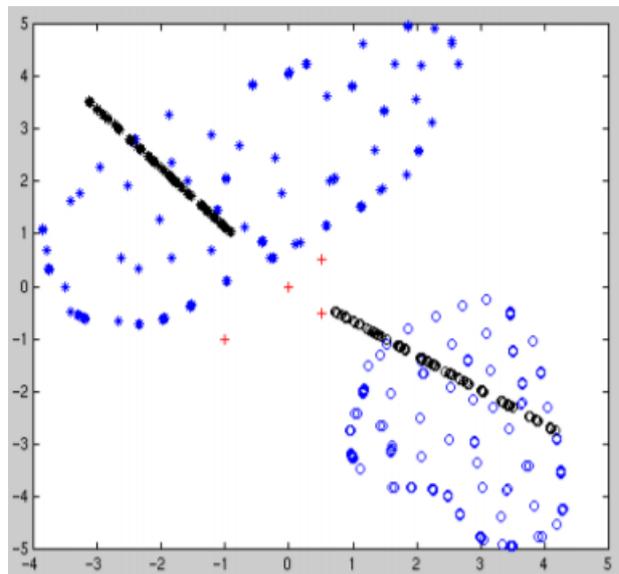


Ilustración 4. Ejemplo de la aplicación del análisis LDA con datos, espacio, puntos de prueba, y región de decisión (Ganapathiraju & Balakrishnama, 1998).

Este tipo de análisis es conveniente utilizarlo, cuando las frecuencias dentro de la clase no son iguales y los desempeños fueron examinados en los datos de prueba generados al azar (Ganapathiraju & Balakrishnama, 1998).

La diferencia principal entre las técnicas de análisis anteriores, se refiere a que PCA realiza la clasificación a nivel de entidades, y LCA realiza la clasificación a nivel de datos. Es importante aclarar que las técnicas anteriores, no son viables de utilizar cuando los datos no son lineales (Nan & Weiwei, 2016); es decir, no funcionarían para no capturar similitudes entre datos de este tipo.

b. PROYECCIÓN NO LINEAR

La desventaja de los métodos de proyección lineal, radica en que son incapaces de encontrar dependencias no lineales entre las variables. Por lo anterior, existen diversos métodos que permiten resolver el problema anteriormente mencionado y algunas de las técnicas existentes en la proyección no lineal son: el escalamiento multidimensional (**MDS**, las siglas en inglés de **Multidimensional Scaling**), isotop, y la incrustación local (**LLE**, por sus siglas en inglés **Locally Linear embedding**). El objetivo de este tipo de técnicas, es para visualizar datos altamente dimensionales como una imagen en 2D.

La técnica **MDS** se encuentra basada en la preservación de las disimilitudes a pares o de las distancias euclidianas. Dentro de la técnica, se proyectan puntos de datos en un plano bidimensional preservando las métricas inter-puntos (Hujun, 2003). La técnica **Isotop**, consiste en un nuevo algoritmo basado en las técnicas de Mapa auto-organizado (**SOM**, siglas en inglés de **Self-Organizing Map**) y el uso de algoritmos como **Sammon**, el cual es implementado en NLM. *“Isotop procede en 3 fases: cuantificación vectorial de los datos brutos, los prototipos vecinos y el mapeo al espacio de proyección”* (Lee & Verleysen, 2002).

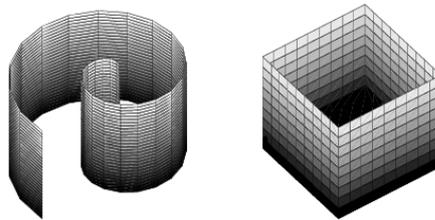


Ilustración 5. Conjunto de datos artificiales: rollo suizo (Swiss roll, izquierda) y caja abierta (open box, derecha) (Hujun, 2003).

En el artículo *“Overview of text Visualization Techniques”*, se hace una comparación (ilustración 6) en donde se aprecia una comparación visual de diferentes métodos o técnicas, en un plano 2D.

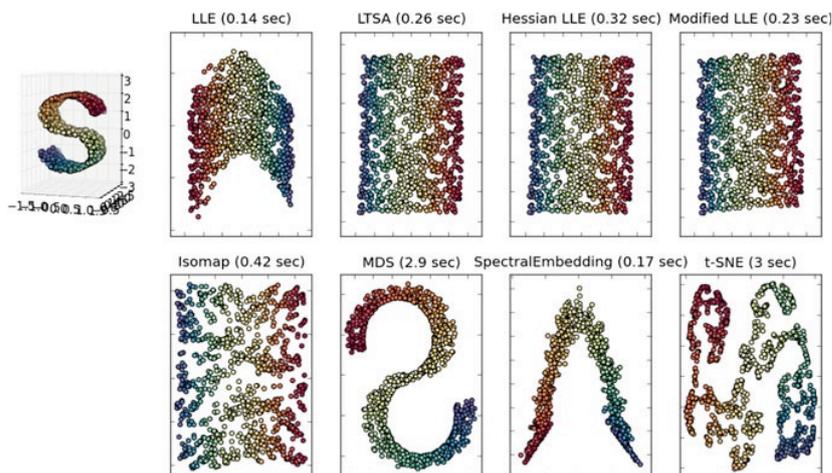


Ilustración 6. Comparación de diferentes técnicas de proyección no lineal (Nan & Weiwei, 2016).

2. TÉCNICAS BASADAS EN GRAFOS

Las técnicas basadas en grafos, permiten la visualización de una gran cantidad de grafos, a través de métodos para transmitir el significado del grafo de manera clara y rápida (Keim D. A., 2000).



Ilustración 7. Clasificación de las técnicas basados en grafos según Daniel A. Keim (Keim D. A., 2000).

Este tipo de técnica permiten establecer relaciones con los datos, independientemente de la dimensionalidad que es tomada en cuenta por otras técnicas, dado que esta puede ser alta o baja. La técnica basada en grafos, permite es aplicable a datos pequeños o de una escalabilidad mediana, así como una dimensionalidad mediana.

Las técnicas basadas en grafos, contemplan los siguientes parámetros (Dias, Yamaguchi, Rabelo, & Franco, 2012):

- Tipo de dato
- Tipo de tarea
- Escalabilidad
- Dimensionalidad
- Posicionamiento de los atributos

a. GRÁFICOS BÁSICOS (BASIC GRAPHS)

Los gráficos básicos como se muestra en la ilustración 15, se pueden dividir en dibujos polilyne, Straight-Line, y Curved-Line.

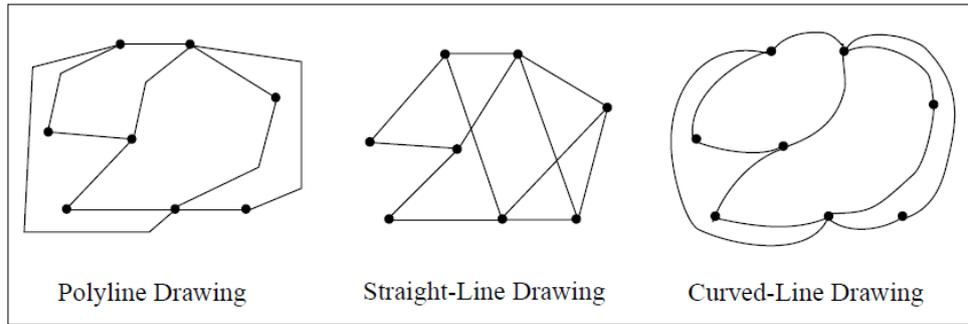


Ilustración 8. Dibujos de gráficos básicos o en 2D, según Daniel A. Keim.

Las propiedades de los dibujos de los gráficos en 2D, según Daniel A. Keim se muestra en la tabla 1.

PROPIEDADES DE LOS GRÁFICOS EN 2D	
Básicas	Planaridad (Planarity)
	Ortogonalidad (Orthogonality)
	Propiedad de Grid (Grid property)
No estéticas (Aesthetics)	Número mínimo de líneas que cruzan.
	Simetrías con visualización óptima.
	Clúster de visualización óptima.
	Número mínimo de "bends" en un gráfico Poyline.
	Distribución uniforme de los vértices.
	uniform edge lengths

Tabla 1. Propiedades d los gráficos en 2D, según Daniel A. Keim (Keim D. A., 2000)

Un ejemplo de la aplicación de este tipo de técnica es el presentado en la ilustración 16, en donde se aprecia la distribución de la información a través de un gráfico que resulta más fácil de entender.

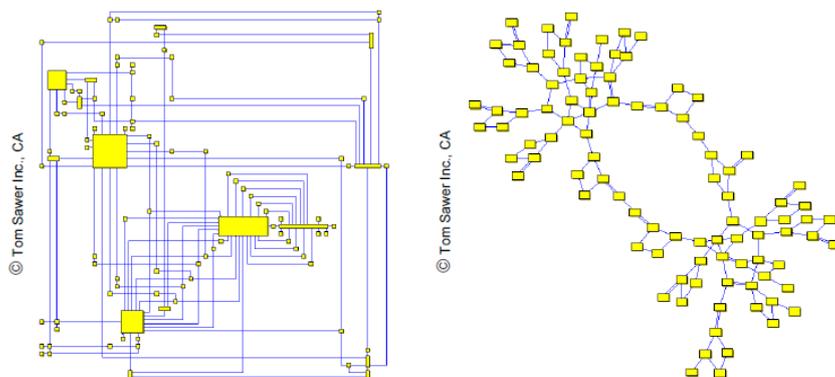


Ilustración 9. Gráfico ortogonal (izquierda) y gráfico simétricamente optimizado (derecha) (Keim D. A., 2000).

b. SPECIFIC GRAPHS

Dentro de esta categoría se encuentra la técnica Hygraphs. Dicha técnica es una estándar que incluye gráficos conteniendo nodos, vínculos (links, en inglés), etiquetas, y contenedores de información. Usando la técnica Hygraphs se desarrolló un sistema más sofisticado denominado Hy+.

El Hypergraph permiten hipertensiones que conectan una gran cantidad de vértices con el grafo, y pueden ser tratados como un subconjunto de vértices, como se aprecia en la ilustración 17. La representación visual de una hipertensión permite distinguir a un conjunto de vértices del resto de vértices (Junghans, 2008).

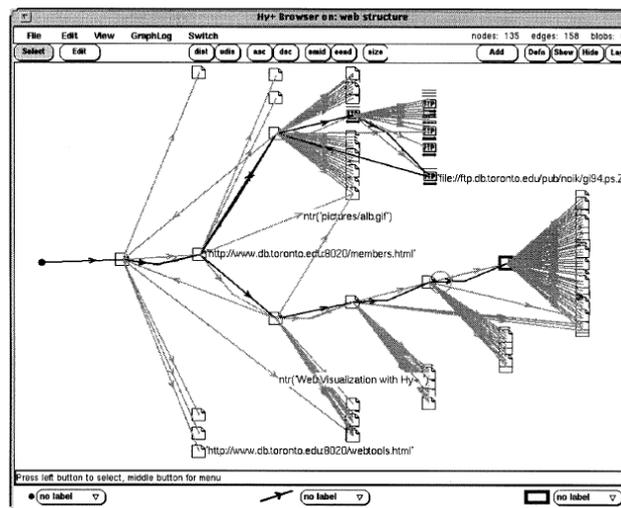


Ilustración 10. Técnica de visualización de información HyGraph (Keim & Kriegel, 1996).

En este contexto existen otras técnicas que permiten trabajar con grafos como lo es "Directed Acyclic" o por sus siglas en inglés DAG (Direct Acyclic Graph); dicha técnica consiste es un gráfico directo y finito que no contiene ciclos. Este tipo de grafos se pueden utilizar para representar subexpresiones comunes en un compilador optimizado, como se aprecia en la ilustración 18.

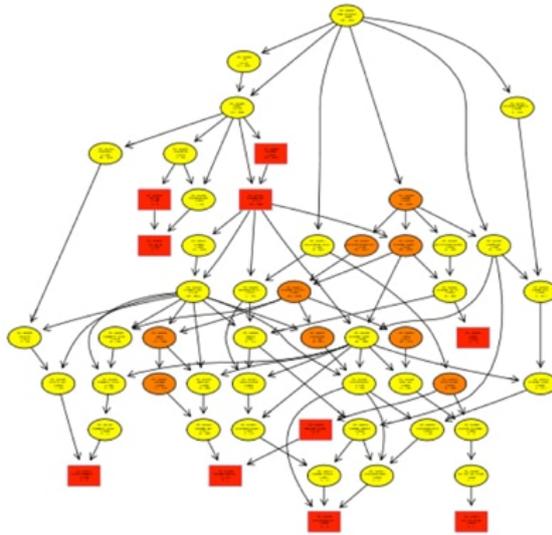


Ilustración 11. Ejemplo de implementación de un DAG. Tomado de Creative Proteomics.

Los grafos DAG están compuestos por una cantidad de nodos y vértices, en donde cada nodo es relacionado con otro. En términos de información, un nodo representa un objeto o una pieza de dato y los vértices directos son vértices que permiten enlazarlos, representando así una relación entre dos nodos. En un diagrama DAG existe el nodo principal y el nodo al que convergen todos (leaf).

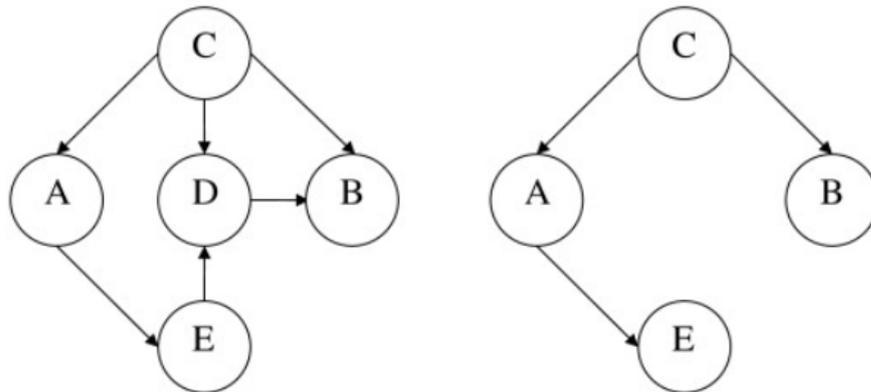


Ilustración 12. Ejemplo de grafos DAG.

3. TÉCNICAS BASADAS EN ICONOS

Existen métodos para mostrar iconos utilizados como medio para expresar los resultados de una exploración de datos. Dentro de este tipo de técnicas para el procesamiento de datos, se encuentran: Little faces, chernoff faces, needle icons, star icons, stick figure icons, color icons; se abordarán dos técnicas de las más relevantes en este estilo.

a. STICK FIGURE ICON

Esta técnica permite mapear datos multidimensionales, en donde dos dimensiones son mapeadas para la visualización, definiendo los ejes (x,y) en la pantalla (screen, término en inglés), y visualizados en figuras de palo (stick figures). Cada figura está conformada por 4 extremidades (limbs, en inglés) y un cuerpo (body, término en inglés).

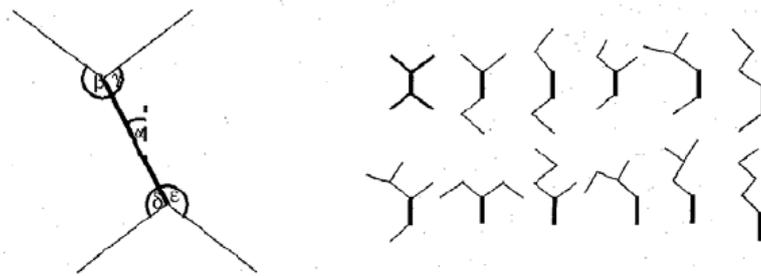


Ilustración 13. Técnica de visualización "Stick Figure", mostrando un icono único y una familia de iconos (Keim D. A., *Visual Techniques for Exploring Databases*, 2000).

Los autores Daniel A. Keim y Hans-Peter Kriegel, mencionan que la técnica de figura de palo fue utilizada para la visualización de los resultados del censo de Estados Unidos de 1980, como se muestra en la ilustración 9 (Keim & Kriegel, 1996).

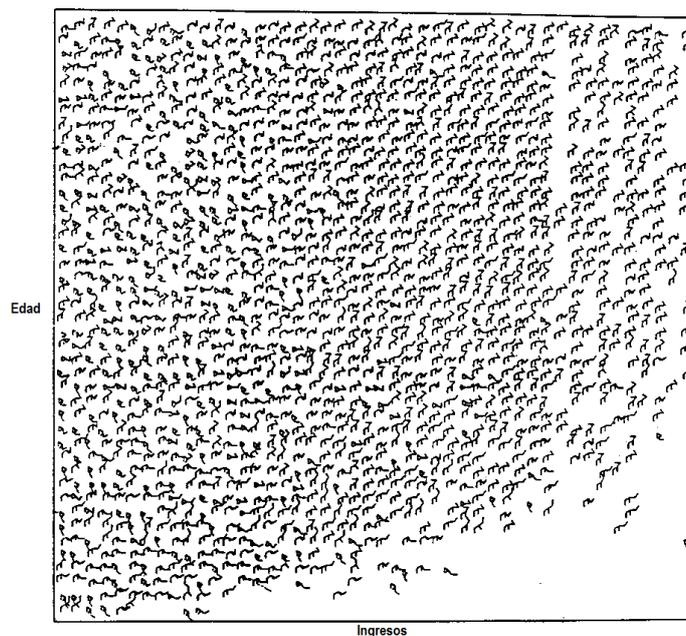


Ilustración 14. Visualización de los datos del censo a través de la técnica Stick Figure. (Keim & Kriegel, *Visualization Techniques for Mining Large Databases: A Comparison*, 1996)

La idea básica de este tipo de técnica es la visualización de los valores de los datos, como características de los iconos; buscando la presentación de todas las variables y los puntos de datos en una sencilla representación.

b. SISTEMAS

Dentro de los sistemas existentes en técnicas de visualización orientadas en grafos se encuentra Tom Sawyer. Dicho software es considerado como un kit de desarrollo completo para la construcción de aplicaciones de visualización de datos a nivel empresarial de alto

desempeño (Tom Sawyer Software, 2017). Dentro de las características principales de la herramienta se encuentran:

- Diagramación avanzada.
- Secciones de animación, barra de progreso, y capacidades multihilo.
- Escalable.
- Sistema avanzado de gestión de eventos.
- Interfaz avanzada de dibujo.

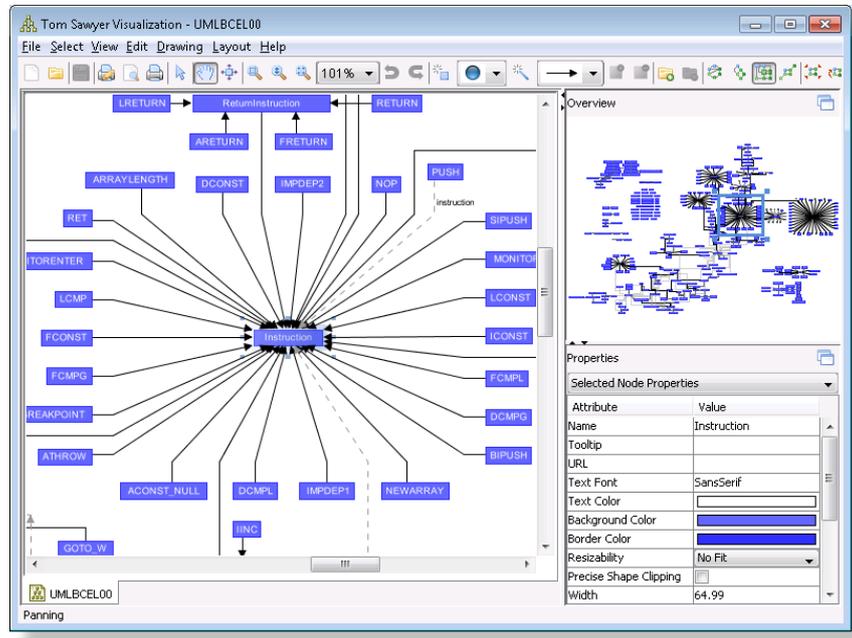


Ilustración 15. Pantalla del software Tom Sawyer.

4. TÉCNICAS BASADAS EN JERARQUÍAS

El objetivo de este tipo de técnicas es aplicar el particionamiento jerárquico del espacio para proporcionar una visualización estructural.



Ilustración 16. Clasificación de las técnicas basadas en jerarquías.

a. DIMENSIONAL STACKING

Esta técnica basada en jerarquía lo que realiza es el particionamiento de un espacio ocupado por un atributo de N dimensiones, en sub-espacios de 2 dimensiones, los cuales son apilados uno sobre otro (Keim D. A., 2000).

Dentro de los puntos que se destacan de la técnica, se encuentran:

- Particionamiento del rango de valores de un atributo en clases.
- Los atributos importantes, deberían ser usados en niveles externos.
- Especial adecuación para datos con atributos ordinales de baja cardinalidad.

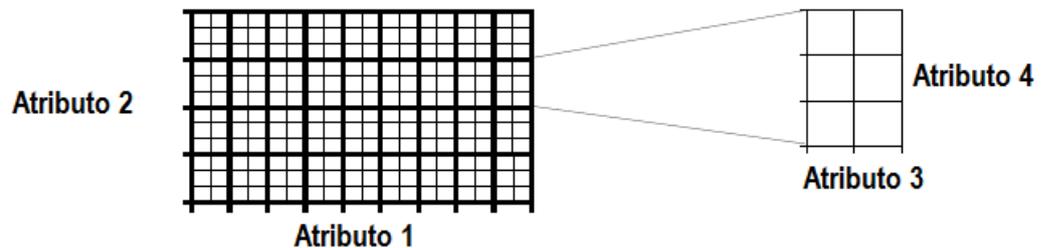


Ilustración 17. Técnica "Dimensional Stacking" (Keim D. A., 2000).

b. TREE MAP

La técnica de visualización denominada *treemap*, utiliza el 100% del espacio disponible en pantalla, mapeando la jerarquía completa en una región rectangular y de cierta manera llenando el espacio completo (Johnson & Shneiderman, 1993). En la ilustración 12 se muestra la visualización de la información a través de un diagrama Venn¹.

¹ Los diagramas Venn, denominados así por su creador John Venn (matemático inglés), proporcionar una vista pictórica de los conjuntos. En un diagrama Venn un rectángulo describe el conjunto y los subconjuntos son representación como círculos (Johnsonbaugh, 2005).

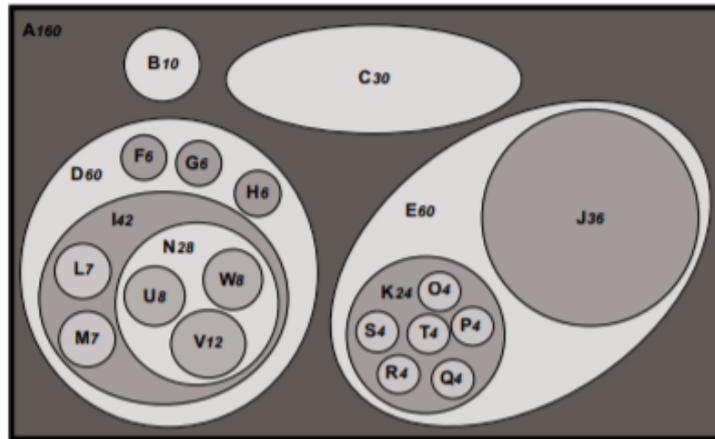


Ilustración 18. Técnica Treemap, representada en un diagrama Venn (Johnson & Shneiderman, 1993).

Lo que realiza en términos generales la técnica *treemap* es:

- Mapea la información jerárquica a una pantalla rectangular 2D.
- Se utiliza el 100% del espacio de la pantalla.
- El control interactivo permite a los usuarios especificar la presentación de la información estructural, así como del contenido.

Dentro de la técnica *treemap*, las dimensiones (x, y) de la pantalla se dividen alternativamente de acuerdo con los valores de atributo, el color de las regiones puede corresponder a un atributo adicional, y los atributos utilizados para el particionamiento y su ordenación serán definidos por el usuario. Esta técnica es adecuada para obtener una visión general de grandes cantidades de datos jerárquicos y para datos con múltiples atributos ordinarios (Keim D. A., 2000).

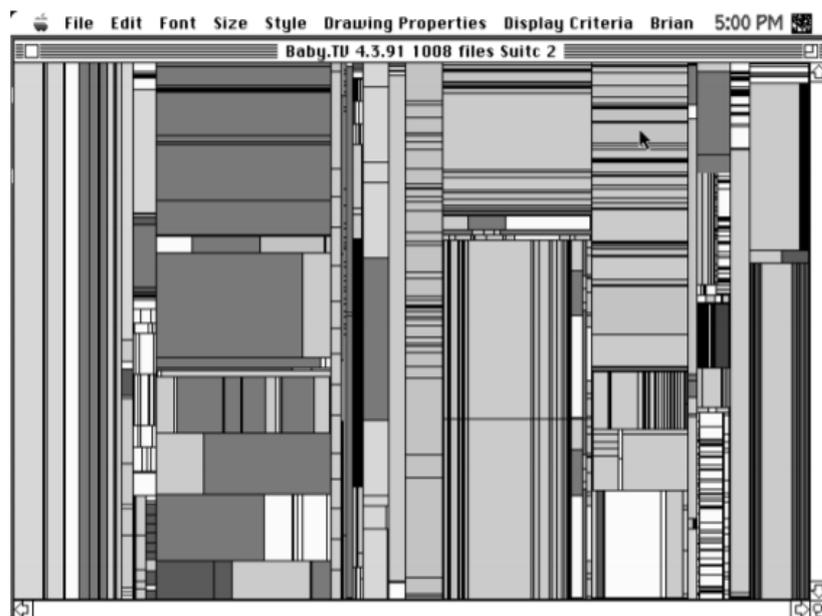


Ilustración 19. Visualización de 1000 archivos, a través de la técnica treemap (Johnson & Shneiderman, 1993).

5. TÉCNICAS BASADAS EN PIXEL

En este tipo de técnica, cada valor del atributo es representado por un píxel de color. Los valores de los atributos, para cada atributo se presentan en sub-ventanas separadas (Keim D. A., Visual Techniques for Exploring Databases, 2000).



Ilustración 20. Clasificación de las técnicas basadas en píxel (Pixel-oriented techniques).

a. TÉCNICAS INDEPENDIENTES DE CONSULTA (QUERY-INDEPENDENT TECHNIQUES)

La técnica de **patrones recursivos** consiste en la generalización recursiva de arreglos iterativos basados en líneas y columnas, en donde se puede especificar el alto (h_i) y el ancho (w_i) para cada nivel de recursividad (Keim D. A., Visual Techniques for Exploring Databases, 2000).

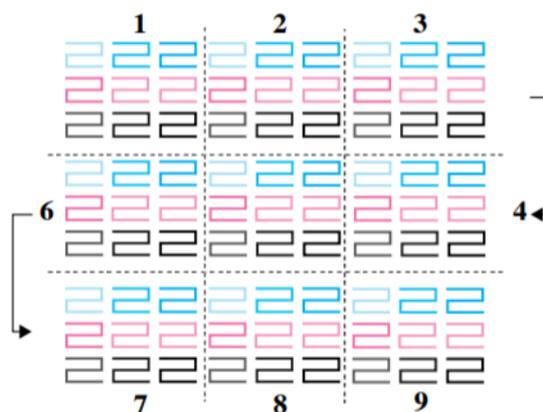


Ilustración 21. Arreglo completamente recursivo (Keim, Kriegel, & Ankerst, Recursive Pattern: A Technique for Visualizing Very Large Amounts of Data, 1995).

La idea básica de la técnica de “patrones recursivos”, es la representación de datos de cualquier cantidad de elementos como sea posible al mismo tiempo, que el número de elementos de datos, solo son limitados por el

número de píxeles de la pantalla (Keim, Kriegel, & Ankerst, Recursive Pattern: A Technique for Visualizing Very Large Amounts of Data, 1995).

b. CHERNOFF FACES

La técnica denominada como **Chernoff faces**, cuyo nombre fue acuñado por Herman Chernoff el cual la propuso en 1973. Lo que propuso fue, un camino para representar datos multivariados de cara (faces, por sus siglas en inglés) de un humano, es decir, cada dato compuesto por n variables se representa mediante una cara, en donde cada variable es asociada a un rasgo de la cara.

Lo interesante de la técnica, radica en la capacidad de los seres humanos de reconocer caras y pequeños cambios que están pueden ocurrir. Los rasgos que se pueden contemplar en el rostro, van desde forma, longitud, ubicación y separación de los ojos, entre otros. En la ilustración 7, se muestra un experimento realizado en la Universidad de Maryland en Baltimore, en donde se analiza la efectividad de la técnica, y presenta las diferentes representaciones según la variabilidad de los datos (Morris, Ebert, & Rheingans, 2000).

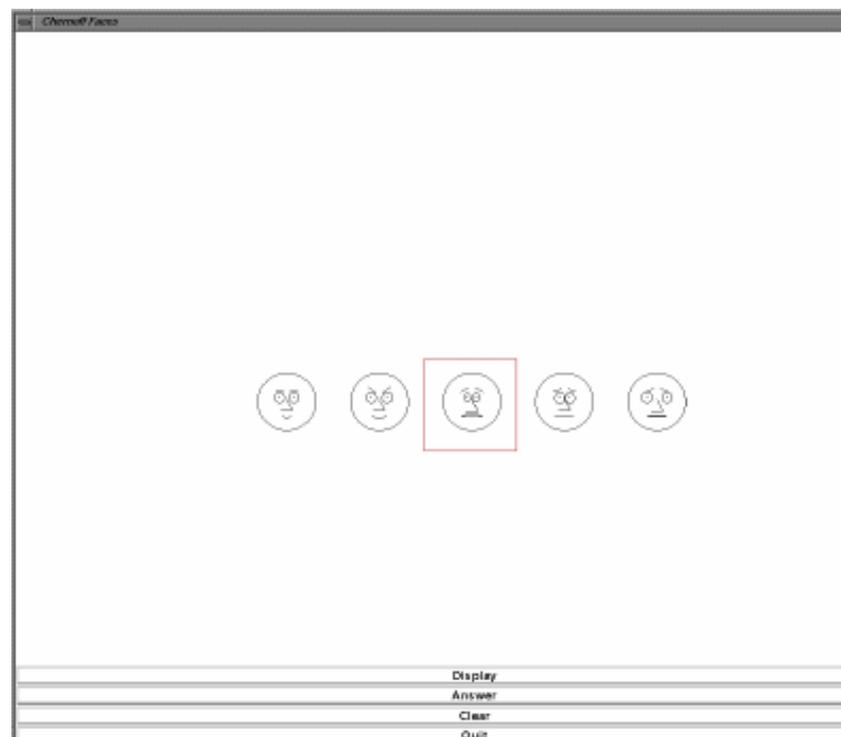


Ilustración 22. Técnica Chernoff Faces. (Morris, Ebert, & Rheingans, 2000)

c. TÉCNICAS DEPENDIENTES DE CONSULTA (QUERY-DEPENDENT TECHNIQUES)

En este tipo de técnicas, los elementos de datos y sus correspondientes consultas son registradas y cada actividad de consulta es simbolizada con

su peso. Dichas actividades son registradas con un pixel de color (Allen & Williams, 2000).

Las ideas básicas de este tipo de técnicas son:

- Elementos de datos y consultas, determinan la distancia.
- Determinar los elementos de datos con las más bajas distancias (totales).
- Mapear distancias a color.
- Extender las distancias con respecto a la distancia total.
- Visualizar el valor de cada distancia, en un pixel de color.

El autor Daniel A. Keim expone el proceso para realizar el cálculo para la visualización de datos, en la técnica dependiente de consulta, como se muestra en la ilustración 23.

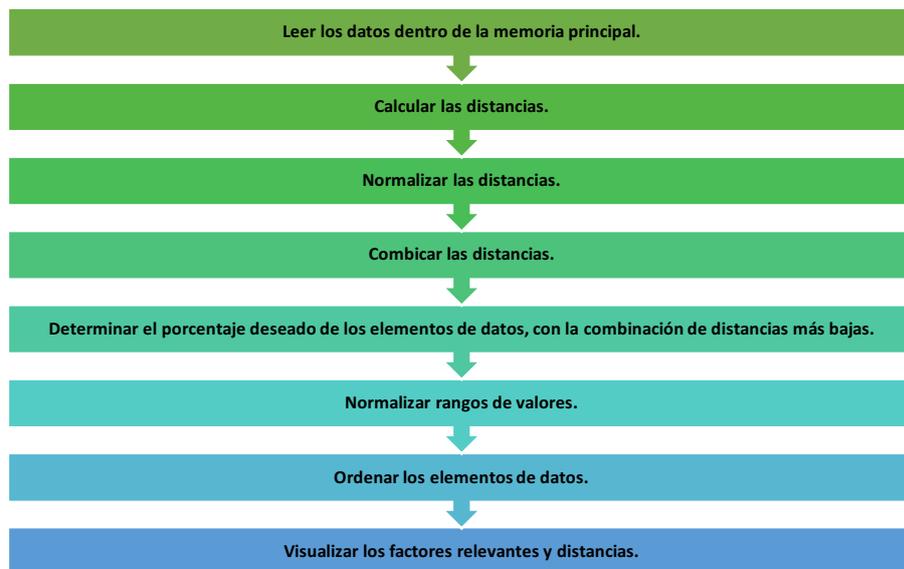


Ilustración 23. Proceso de cálculo para visualizaciones (Keim D. A., *Visual Techniques for Exploring Databases*, 2000).

La técnica denominada espiral (**Spiral Technique**, término en inglés), consiste de consultas que son agrupadas en forma de espiral, en función de las actividades generales de la consulta. Daniel A. Keim, afirma que la técnica de espiral cumple los siguientes puntos:

- Los valores para cada uno de los atributos, son presentados en sub-ventanas separadas.
- La agrupación dentro de las sub-ventanas es acorde al total global de la distancia.

Visualización de un dato de seis dimensiones.

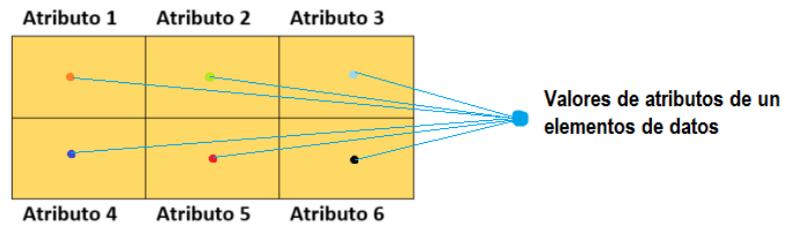


Ilustración 24. Técnica de espiral (Spiral Technique) (Keim D. A., *Visual Techniques for Exploring Databases*, 2000).

Otra variante de la técnica de espiral es la de serpiente-espiral (snake-spiral technique); la idea para visualizar los datos en la pantalla es presentar los datos correctos en el centro de la pantalla y los datos que son aproximados, serán ordenados de acuerdo a su distancia total (o relevancia) en forma de serpiente-espiral alrededor de la región, como se muestra en la ilustración 25.

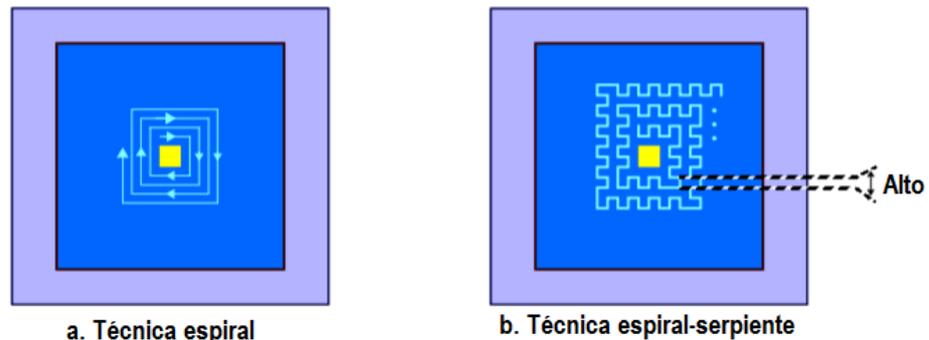
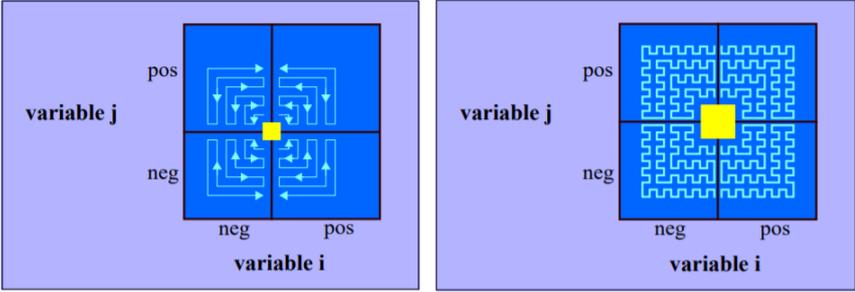


Ilustración 25. La técnica de serpiente-espiral. (Keim D. A., *Pixel-oriented Visualization Techniques for Exploring Very Large Databases*, 1996)

La técnica de ejes(axes), permite mejorar la técnica espiral (spiral technique) al incluir alguna retroalimentación sobre la dirección de la distancia en la visualización. Dentro de las ventajas de esta técnica se encuentran (Keim D. A., *Pixel-oriented Visualization Techniques for Exploring Very Large Databases*, 1996):

- Asignación de dos variables a los ejes(axes).
- Ordenar los datos con base en la dirección de la distancia.
- En caso de una variable de distancia con datos negativos, estos se disponen a la izquierda. Los datos positivos se disponen a la derecha. Para el resto de las variables, las negativas se ubican en la parte inferior y las positivas en la parte superior (ilustración 26).
- Los datos se dividen en cuatro subconjuntos, proporcionando de esta manera información adicional sobre la posición de los datos (con respecto a las variables asignadas a los ejes).
- El número de datos que pueden visualizarse dentro de los cuadrantes es ligeramente inferior.



a. Técnica de ejes (Axes Technique)

b. Técnica de serpiente-ejes (Snake-Axes Technique)

Ilustración 26. Posicionamiento de los datos en la técnica de ejes (Keim D. A., Pixel-oriented Visualization Techniques for Exploring Very Large Databases, 1996).

CONCLUSIÓN

De las técnicas mencionadas anteriormente, existen aquellas que no son empleadas debido a distintas circunstancias. En la ilustración 27 se presentan aquellas técnicas que no fueron empleadas por herramientas de visualización de código abierto (open source, término en inglés), según un estudio realizado por Ana Cristina Umaquina, *et al* (Umaquina, Peluffo, & Alvarado, 2016).

Jerarquías	Dimensional Stacking
	Worlds-within-Worlds
	Cone Trees
	InfoCube
Gráficos	Straight-Line, Polyline
	DAG y Symmetric
	Hy+, Tom Sawyer, SeeNet, Narcissus
Híbridas	Hybrid
RadViz	RadViz
Tree	Tree
	Radial
	Ballon
	Icile
Interactive	Dinamics
	Demand
Visual Structures	1,2,3,Dimensions
	Multi-d, Multidimensional
	Tree and networks
Others	Relationship complete data

Ilustración 27. Herramientas de visualización no empleadas según Ana Cristina Umaquina, et al.

Las técnicas para la visualización de información, permiten procesar y presentar millones de datos, los cuales no son posibles de procesar con herramientas convencionales para la generación de reportes. Actualmente, cada vez los sistemas generan una gran cantidad de información, las cuales pueden ser trabajadas en distintas dimensiones, si se empieza a trabajar con herramientas de Big Data.

En el presente documento no se emite un juicio sobre que técnica es la mejor sobre otra, dado que, no existe un criterio o estándar que permita evaluar este tipo de herramientas. Cada usuario debe determinar que técnica es la más adecuada, según las necesidades y requerimientos de un proyecto o sistema. Es importante destacar que cada técnica presenta opciones de representación visual, en lo cual pueden determinar cuál es mejor que otra. Las técnicas de visualización permiten determinar de un trabajo de investigación, que parámetros son más utilizados y requeridos por el usuario. En la ilustración 28 se muestran las herramientas más utilizados según Ana Cristina Umaquina, *et al* (Umaquina, Peluffo, & Alvarado, 2016).

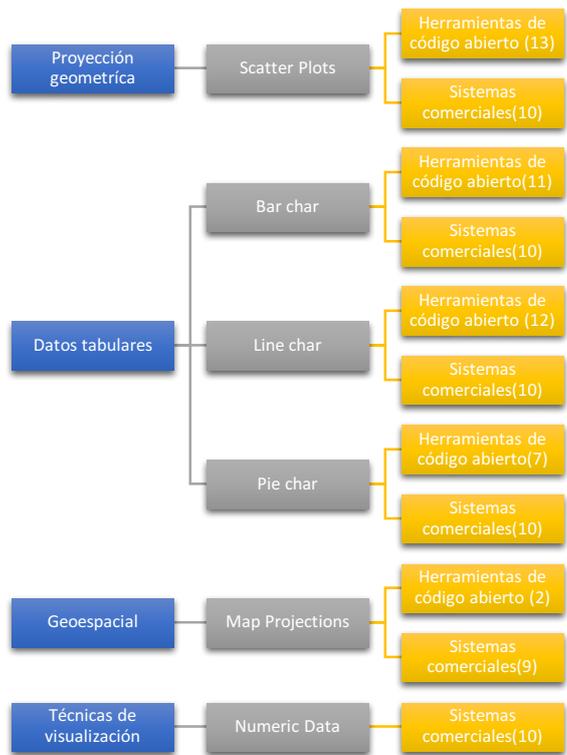


Ilustración 28. Técnicas de visualización de Big Data más utilizadas, según Ana Cristina Umaquina, et al (Umaquina, Peluffo, & Alvarado, 2016).

BIBLIOGRAFÍA

- Allen, K., & Williams, J. G. (2000). *Encyclopedia of Computer Science and Technology* (ilustrada ed., Vol. 42). USA: CRC Press.
- Dias, M. M., Yamaguchi, J., Rabelo, E., & Franco, C. (2012). Visualization Techniques: Which is the Most Appropriate in the Process of Knowledge Discovery in Data Base? En A. Karahoca, *Advances in Data Mining Knowledge Discovery and Applications* (págs. 155-180). USA: INTECH.
- Ganapathiraju, A., & Balakrishnama, S. (1998). *LINEAR DISCRIMINANT ANALYSIS - A BRIEF TUTORIAL*. Institute for Signal and Information Processing. Mississippi: Mississippi State University.
- Hujun, Y. (2003). *Nonlinear Multidimensional Data Projection and Visualisation*. Dept. of Electrical Engineering and Electronics. Manchester: University of Manchester Institute of Science and Technology .
- Johnson, B., & Shneiderman, B. (1993). *Sparks of Innovation in Human-Computer Interaction*. Norwood: Ablex .
- Johnsonbaugh, R. (2005). *Matemáticas discretas*. México: Pearson Educación.
- Junghans, M. (2008). *Visualization of Hyperedges in Fixed Graph Layouts*. Brandenburg: Brandenburg University of Technology Cottbus.
- Keim, D. A. (1996). Pixel-oriented Visualization Techniques for Exploring Very Large Databases. *Journal of computational and graphical statistics*, 58-77.
- Keim, D. A. (2000). *Visual Techniques for Exploring Databases*. Institute for Computer Science. Wittenberg: University of Halle.
- Keim, D. A., & Kriegel, H.-P. (December de 1996). Visualization Techniques for Mining Large Databases: A Comparison. *IEEE Transactions on Knowledge & Data Engineering*, 8, 923-938. doi:10.1109/69.553159
- Keim, D. A., Kriegel, H.-P., & Ankerst, M. (1995). *Recursive Pattern: A Technique for Visualizing Very Large Amounts of Data*. Institute for Computer Science, University of Munich. Atlanta: Proc. Visualization.
- Lee, J. A., & Verleysen, M. (Agosto de 2002). Nonlinear Projection with the Isotop Method. (J. Dorronsoro, Ed.) *ICANN'2002 proceedings - International Conference on Artificial Neural Networks*, 933-938.

- Morris, C. J., Ebert, D. S., & Rheingans, P. (Mayo de 2000). An Experimental Analysis of the Effectiveness of Features in Chernoff Faces. *PROC SPIE INT SOC OPT ENG*, 3905, 12-17.
- Nan, C., & Weiwei, C. (2016). Overview of Text Visualization Techniques. En *Introduction to Text Visualization* (1 ed., págs. 11-40). Atlantis Press. doi:10.2991/978-94-6239-186-4
- Scholz, M. (2006). *Approaches to analyse and interpret biological profile data*. University of Potsdam. Germany: Postdam.
- Tascón, M. (2013). Introducción: Big Data (Pasado, presente y futuro). *Telos* 95, 48-51.
- Tom Sawyer Software. (01 de 01 de 2017). *Tom Sawyer Software*. Obtenido de Tom Sawyer Visualization: <https://www.tomsawyer.com/products/visualization/>
- Umaquina, A. C., Peluffo, D. H., & Alvarado, J. C. (2016). *Estudio descriptivo de técnicas aplicadas en herramientas Open Source y comerciales para visualización de información de Big Data*. Ecuador: Corporación Universitaria Autónoma de Nariño – Universidad de Salamanca.