

1

Introducción

En esta tesis se presenta un esquema de visualización interactiva para la exploración de grandes colecciones digitales, de forma que al inicio se brinde a los investigadores un panorama general de las mismas y posteriormente ellos puedan introducirse poco a poco a elementos más específicos.

La visualización se basa en un modelo jerárquico conocido como *Sunburst* [Stasko & Zhang 2000], el cual se modifica mediante la introducción de nuevos elementos interactivos como lo son: filtros que delimitan los resultados a sólo aquellos que cumplan con ciertas características establecidas por el usuario; un árbol de navegación para mantener el contexto; entre otros. Todo esto tiene la finalidad de ayudar en actividades de investigación.

1.1 Definición del problema

En la actualidad existe gran cantidad de información sobre diferentes áreas de estudio en forma de libros, escritos y documentos tanto físicos como digitales que se guardan en diferentes instituciones. Gracias a la llegada de internet, estas instituciones han creado repositorios que permiten al usuario acceder, vía web, al conocimiento que generan. Aunque al momento de realizar consultas es complicado explorar los resultados debido a la vasta cantidad de texto que se presenta, impidiendo llegar a la plena comprensión y conocimiento del repositorio.

Para lo anterior se han creado dos tipos de soluciones generales. Una es la que ofrece un listado de documentos acompañados con algunas sugerencias de escritos similares en respuesta a términos o palabras clave proporcionadas por el usuario. Sin embargo, esto puede llegar a llenar la pantalla con gran cantidad de texto, saturando de información al usuario y, en consecuencia, éste puede no lograr enfocarse en los datos que necesita. También es importante mencionar que esta solución no ofrece una perspectiva global, por lo que con frecuencia no se llega a comprender la esencia de la colección ni las pequeñas partes que la conforman, desaprovechando en su totalidad estos recursos [Chang, Leggett, Furuta, Kerne, Williams, Burns & Bias 2004].

La otra solución son visualizaciones que resuelven los dos problemas previos con el uso de interfaces que contienen formas y colores que representan aspectos como jerarquías y relaciones entre documentos. Al mostrar la información a través de gráficos, este tipo de recurso evita saturar la pantalla de texto, simplificando la exploración y búsqueda, además de que al mismo tiempo logra brindar una buena perspectiva de los repositorios gracias a su naturaleza visual. La problemática que se presenta en este caso es que con frecuencia no se cubren las necesidades del usuario al momento de que éste desea delimitar la cantidad de datos que se observan, es decir, a pesar del uso de formas y colores para facilitar la búsqueda, se requiere una forma de restringir la exploración a elementos que sólo cubran con ciertas características y requerimientos. Esto último sirve para reducir la información a sólo aquella potencialmente útil, logrando sacar más provecho de la búsqueda, exploración y análisis de datos. Muchos trabajos existentes no utilizan este recurso para el manejo de grandes cantidades de información y se habla de ellos en el Capítulo 2 por ser una motivación para la creación del presente proyecto.

1.2 Hipótesis

Este proyecto busca explorar los beneficios de los esquemas de filtrado como apoyo de una visualización jerárquica tipo *Sunburst* para la exploración en grandes colecciones de documentos de todo tipo que se encuentran en repositorios institucionales. Al mismo tiempo, se pretende llegar también al diseño e implementación de un modelo conformado por los componentes ya mencionados, para así desarrollar así un prototipo de visualización de información usable y funcional llamado *SunGroups*.

1.3 Solución propuesta

La idea central de esta tesis es mejorar las interfaces existentes que ayudan en la exploración de grandes cantidades de información, enfocándose inicialmente en repositorios institucionales. Para lograr esto, se buscó un estilo de visualización de información adecuado y sobre todo, se trabajó en un esquema de filtrado [Ahlberg & Shneiderman 1994] eficiente que se convirtiera en un apoyo para el usuario al momento de utilizar la visualización.

Primeramente se encontró que una de las características que cuenta con grandes beneficios es la jerarquización, ya que este tipo de estructura es muy abundante en el mundo real tanto en la naturaleza como en la sociedad, convirtiéndose así en un concepto con el cual los usuarios están muy familiarizados y, por lo tanto, les es sencillo manejar y entender las representaciones que cuentan con esta particularidad. Otra ventaja de la estructura jerárquica es que con ella se logra abordar un problema al descomponer los elementos complejos en más simples y sencillos para de ese modo entender el conjunto [Sáez 2009]. Gracias a este hallazgo el estudio se centró en esquemas que utilizaran jerarquía, llegando a dos visualizaciones que cuentan con buenas referencias y que son de utilidad: *TreeMap* [Johnson & Shneiderman 1991] y *Sunburst* [Stasko & Zhang 2000]. La primera consiste en la muestra de datos jerárquicos mediante rectángulos anidados. La segunda deriva de la idea de la primera, utilizando un plano circular en lugar de uno rectangular. A detalle, *Sunburst* se trata de un plano radial que marca jerarquías a través de círculos concéntricos que a su vez se dividen en sectores que indican campos.

El que *Sunburst* sea ocupado para la visualización de diferentes tipos de datos, por ejemplo: análisis de muros de fuego [Mansmann, Göbel, & Cheswick 2012], representación de las palabras importantes de un libro [Collins, Carpendale, & Penn 2009], entre otras cosas; demuestra que cuenta con elementos eficientes y ventajas que se pueden aprovechar. Por ello se decidió utilizar este esquema. Posteriormente se observó que en *Sunburst* entre más círculos concéntricos y sectores existen, el espacio para el contenido textual dentro de éstos se reduce, haciendo que las palabras se corten o el tamaño de la fuente disminuya, dificultando la lectura. Para enfrentar este obstáculo se opta por manejar tres niveles máximos en el *Sunburst*, gráficamente hablando. Obviamente esto conlleva a que si lo que se desea representar contiene más de tres niveles, del cuarto en adelante no se observarán en primera instancia. Esto implica recurrir a lo que en esta tesis se le denomina “profundidad controlada”, la cual consiste en que al dar clic en algún sector, éste toma el lugar del nivel central del *Sunburst* y todos sus subsectores se colocan de acuerdo a la jerarquía a la que pertenecen. Como los niveles anteriores al seleccionado desaparecen de la vista, queda espacio para agregar los subsectores del nivel que se ha convertido en el principal o de aquellos que son sus subordinados. Por supuesto, respetando siempre la idea de tres niveles máximo. Sin embargo, es necesario un componente que informe al usuario la ruta que va siguiendo, es decir, a lo que va dándole clic. Este complemento es un árbol de navegación que permite mantener el contexto al dar a conocer los niveles jerárquicamente superiores a los mostrados en pantalla, además de que el árbol también funciona como enlaces a niveles superiores. Todo lo detallado se le conoce en este escrito como “navegación contextual”. Esta combinación de árbol de navegación y *Sunburst* se elige porque recibe buenas críticas en la interfaz de muros de fuego y también porque, gracias a ella, se mantiene la idea de un máximo de tres niveles para una mayor legibilidad.

Aunque lo anterior funciona como un filtro al reunir los documentos en grupos, se contempla agregar un esquema de filtrado más específico que concede la posibilidad de manipular una interfaz para analizar sólo las publicaciones que cumplan con requerimientos más precisos, ya que así se le ofrece al usuario lo esencial. Todo este conjunto, en resumen, ayuda a clasificar y delimitar la cantidad de datos que se observan.

Por último, se añade al concepto de *Sunburst* un código de color en degradados para enfatizar las jerarquías, pero dando la opción de cambiar dicha paleta de colores a otra que agrade visualmente al usuario.

En conclusión, este proyecto propone un esquema de visualización, basado en los conceptos de *Sunburst*, “profundidad controlada”, “navegación contextual”, filtros y códigos de color, buscando así cubrir las necesidades de un usuario que se dedica a explorar dentro de repositorios institucionales.

1.4 Contribuciones

Uno de los objetivos de un proyecto de tesis es brindar contribuciones al ámbito de la investigación y, es por ello que debe realizarse todo un proceso de búsqueda e indagación para lograr plantear ideas que aporten algo nuevo. Para este proyecto se hace dicha tarea, dando a conocer el análisis resultante en capítulos siguientes. De esta manera se establecieron inicialmente las principales contribuciones teóricas que ofrece *SunGroups*:

- a) La primera es el uso de una visualización tipo *Sunburst* especialmente para la exploración dentro de repositorios institucionales, porque visualizaciones de esta índole se centran en archivos, palabras clave de documentos o en el análisis de muros de fuego. Esto da pie a que idear nuevas formas de implementar *Sunburst* para la visualización del contenido de repositorios institucionales.
- b) En *SunGroups* las publicaciones se agrupan tomando en cuenta su contenido, es decir, se toman las palabras clave y, si una serie de publicaciones las contienen, se colocan en un mismo grupo. Cuando un grupo contiene demasiadas publicaciones, éste se subdivide siguiendo la mecánica de las palabras claves ya mencionadas. Este proceso continúa hasta que un grupo contenga un número manejable de documentos. De esta forma, regresando a la idea de la visualización tipo *Sunburst*, en *SunGroups* al darle clic a algún subnivel, este toma el lugar del nivel principal y aparecen, según sea el caso, uno o dos niveles de jerarquía menor. Esto se repite hasta llegar al grupo final que contiene las publicaciones y así mostrarlas en otra área de la interfaz. Esta aportación, que es una de las más novedosa, mejora la legibilidad del contenido de los sectores de los círculos ya que el manejar pocos niveles provee de más espacio para la escritura. A todo esto se le

ha llamando “profundidad controlada” y trabaja de como un filtro para ayudar al usuario a enfocarse en lo que necesita. Este concepto brinda una nueva forma de utilizar el esquema *Sunburst* para la representación de grandes cantidades de información.

- c) Se introduce un árbol de navegación para complementar la “profundidad controlada”, así se mantiene el contexto al conocer los grupos a los cuales se ha entrado, es decir, el camino que se ha seguido para llegar al esquema actual y conocer sus niveles jerárquicamente superiores. Además existe la “navegación contextual” porque el nombre de los grupos en el árbol de navegación funcionan como enlaces a los mismos.
- d) Para delimitar aún más los resultados y ofrecer los que van más acorde con lo requerido por el usuario, se añaden a la interfaz filtros más específicos, por ejemplo: de tiempo, ubicación, etc. Cabe resaltar que la combinación de *Sunburst* y filtros para la exploración dentro de repositorios institucionales es inexistente en trabajos realizados hasta ahora, por lo que al aplicarla a una interfaz y comprobar su eficacia, puede llegar a aplicarse en otro tipo de colecciones o interfaces.
- e) Por último, otra aportación es el manejo de un código de color que enfatiza las jerarquías con el uso de degradados. El color oscuro marca niveles superiores y al aclararse la jerarquía va disminuyendo. Aún así se tiene la posibilidad de cambiarlo en caso de que el usuario prefiera otros tonos. Al comprobar la utilidad del código de color utilizado se obtiene una herramienta más para facilitar el entendimiento de visualizaciones jerárquicas. También al dar la posibilidad de cambiar los tonos, el público meta puede extenderse con aquellos que sufran alguna *CVD* (*Color Vision Deficiency* - Deficiencia en la visión del color) [Flatla & Gutwin 2010].

Como añadido, la contribución práctica que ofrece *SunGroups* es el uso de la librería *Data Driven Document (D3)* para la visualización de colecciones dentro de repositorios institucionales, con lo que se probará si este tipo de herramienta puede ser aplicado para el uso de interfaces que busquen mostrar grandes cantidades de información.

1.5 Organización del documento

Esta tesis se compone por seis capítulos, dentro de los cuáles se incluye esta introducción. A continuación se describe brevemente de que trata cada uno.

El **Capítulo 2** ofrece una visión general del proyecto al dar a conocer conceptos básicos. También se presentan los trabajos más representativos en relación con la visualización de información, esquemas de filtrado y otros elementos que se utilizaron para el presente proyecto. Particularmente en este capítulo se busca el análisis de trabajos existentes relacionados con lo que se propone en esta tesis para realizar una comparación y así, contextualizar las aportaciones ofrecidas por *SunGroups*.

El **Capítulo 3** explica el proceso que se siguió para desarrollar el concepto del *software* llamado *SunGroups*, para posteriormente dar a conocer a detalle sus aportaciones.

El **Capítulo 4** da a conocer las herramientas y la arquitectura empleadas para el desarrollo de un prototipo de *SunGroups*. De igual forma se da una breve explicación del sistema de agrupación utilizado para la visualización tipo *Sunburst*. Finalmente, se describen los elementos gráficos y el funcionamiento del prototipo final.

El **Capítulo 5** muestra los resultados de las evaluaciones con usuarios que se realizaron para determinar la utilidad y el potencial del *software*.

El **Capítulo 6** ofrece las conclusiones a las que se llegaron gracias a la investigación realizada en esta tesis y a la implementación del proyecto planteado.